Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
Department of Psychology, University of Michigan
Ann Arbor, Michigan

DAVID A. GRANT, Associate Editor
University of Wisconsin

WILLIAM K. ESTES, Associate Editor
Indiana University

Consulting Editors

NORMAN H. ANDERSON, *University of California, Los Angeles*
E. JAMES ARCHER, *University of Wisconsin*
FRED ATTNEAVE, III, *University of Oregon*
CLETUS J. BURKE, *Indiana University*
JAMES DEESE, *Johns Hopkins University*
PAUL M. FITTS, *University of Michigan*
FRANK A. GELDARD, *University of Virginia*
JAMES J. GIBSON, *Cornell University*
HAROLD W. HAKE, *University of Illinois*
ARTHUR L. IRION, *Tulane University*
HOWARD H. KENDLER, *New York University*
HERSCHEL W. LEIBOWITZ, *International Business Machines Corporation, Bethesda, Maryland*
DONALD B. LINDSLEY, *University of California, Los Angeles*
KENNETH MACCORQUODALE, *University of Minnesota*
QUINN MCNEMAR, *Stanford University*
LEO POSTMAN, *University of California, Berkeley*
L. STARLING REID, *University of Virginia*
KENNETH W. SPENCE, *State University of Iowa*
DELOS D. WICKENS, *Ohio State University*

CONTENTS

The Partial Reinforcement Effect Sustained through Blocks of Continuous Reinforcement: J. THEIOS 1

The Effects of Language Habits on the Acquisition and Retention of Verbal Associations: L. POSTMAN 7

Conditioned Diminution of the Unconditioned Response as a Function of the Number of Reinforcements: H. D. KIMMEL AND H. S. PENNYPACKER 20

The Effects of Drive and Discrimination Training on Stimulus Generalization: D. R. THOMAS 24

The Learning of Responses to Multiple Weighted Cues: S. A. SUMMERS 29

Stimulus-Response Contiguity in Classical Aversive Conditioning: R. A. CHAMPION 35

Associations, Sets, and the Solution of Word Problems: M. A. SAFREN 40

Sequential Interferences Demonstrated by Serial Reconstructions: E. B. COLEMAN 46

The Effect of Order of Approximation to the Statistical Structure of English on the Emission of Verbal Responses: K. SALZINGER, S. PORTNOY, AND R. S. FELDMAN 52

The Effects of Reward and Knowledge of Results on the Performance of a Simple Vigilance Task: R. R. SIPOWICZ, J. R. WARE, AND R. A. BAKER 58

Some Conditions Influencing the Acquisition and Utilization of Cues: L. S. BRALEY 62

Effect of Stimulus Condition and Reaction Time Information on Spatial Stimulus Generalization: C. Y. NAKAMURA AND J. W. KASWAN 67

Stimulus Generalization as a Function of the Frame of Reference: D. R. THOMAS AND C. G. JONES 77

Effect of Reward Magnitude, Percentage of Reinforcement, and Training Method on Acquisition and Reversal in a T Maze: W. F. HILL, J. W. COTTON, AND K. N. CLAYTON 81

Paired-Associate Learning under Simultaneous Repetition and Nonrepetition Conditions: W. F. BATTIG 87

Supplementary Report: Effects of Stimulus Association Value and Exposure Duration on R-S Learning: N. CASSEM AND D. H. KAUSLER 94

Supplementary Report: Effects of Instructions on Extinction and Recovery of a Conditioned Avoidance Response: K. E. MOYER AND R. H. LINDLEY 95

American Psychological Association

Vol. 64, No. 1 July 1962

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc., at Prince and Lemon Sts., Lancaster, Pa. The subscription rate per volume is $10.00, or $20.00 annually. Single copies are $2.00. Subscriptions, orders, address changes, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This Journal publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects, except when the latter are specifically oriented toward the extension of behavior theory. Experimental psychometric studies and studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology.

Normally, articles exceeding 20 printed pages cannot be accepted. Within this limit, the piecemeal experiment-by-experiment reporting of psychological research is discouraged, and the reporting of the data and conclusions of substantial segments of research efforts is encouraged. Specifically, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) *must* be presented in a single article. Special provision is made for the publication of Supplementary Reports and Replication Reports in articles of not more than 1.5 printed pages (see Editor's Note, this Journal, 1957, **53**, 1–2).

Manuscripts *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc., as described in the *Publication Manual of the American Psychological Association*. When in doubt about practices of the Journal, authors should examine a recent issue.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1962 the publication cost is $30 per printed page.

All **regular** articles, **early publication** articles, **supplementary reports,** and **replication reports** should be submitted to the Editor-elect: David A. Grant: Department of Psychology, University of Wisconsin; Madison 6, Wisconsin. Manuscripts must be accompanied by an abstract of 100 to 120 words typed on a separate sheet of paper. The abstract should conform to the style of *Psychological Abstracts*. Detailed instructions for preparation of the abstracts may be obtained from the Editor or from the APA Central Office.

HELEN ORR
Managing Editor

RUTH P. JAGOLINZER
Editorial Assistant

Second-class postage paid at Lancaster, Pa.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 10th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

© 1962 by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 64, NO. 1 JULY 1962

THE PARTIAL REINFORCEMENT EFFECT SUSTAINED THROUGH BLOCKS OF CONTINUOUS REINFORCEMENT [1]

JOHN THEIOS [2]

Stanford University

The discrimination interpretation of partial reinforcement (Bitterman, Fedderson, & Tyler, 1953; Mowrer & Jones, 1945) holds that resistance to extinction is a function of the degree of similarity between the training reinforcement schedule and the continuous nonreinforcement of extinction. According to this point of view, it is relatively easy for continuously reinforced *S*s to discriminate when extinction starts because of the sharp contrast in the reinforcing events at the onset of extinction. Partially reinforced *S*s have difficulty in discriminating the transition from training to extinction because they have experienced runs of nonreinforced trials, and the first few extinction trials provide a stimulus situation which is similar to that experienced during the nonreinforced runs in training.

In stimulus generalization terms, continuously reinforced *S*s experience a large stimulus change during extinction which results in a large decrement in response strength. Partially reinforced *S*s experience only a small stimulus change, and their response strength is decreased only slightly. The difference in resistance to extinction is attributable to the differential stimulus generalization decrements.

A possible test of the discrimination hypothesis would consist of varying the number of continuously reinforced trials interpolated between an initial partial reinforcement series and an extinction series. The continuously reinforced trials would serve to isolate the partial reinforcement training from extinction by permitting the partially reinforced *S*s to have a sharp contrast in the reinforcing events at the onset of extinction similar to that experienced by con-

[1] This study is based on a dissertation submitted to the faculty of Stanford University in partial fulfillment of the requirements for the PhD degree. The writer wishes to express his gratitude to the chairman of his dissertation committee, Douglas H. Lawrence, and to Gordon H. Bower for their cogent advice and generous aid during all stages of the development of this study. Thanks are also due to Ernest R. Hilgard and Patrick Suppes for their valuable comments during the later stages. The research was conducted while the writer was a National Science Foundation Cooperative Graduate Fellow.

[2] Now at the University of Texas.

tinuously reinforced *Ss*. This stimulus change experienced by *Ss* that have received some continuous reinforcement following partial reinforcement should be larger than the change experienced by *Ss* that received only partial reinforcement, and according to the discrimination hypothesis, the former *Ss* should extinguish faster than the latter *Ss*. In general, the longer the block of continuous reinforcement given after partial reinforcement, the less similarity there should be between training and extinction. Therefore, resistance to extinction is expected to decrease with the number of continuously reinforced trials interpolated between partial reinforcement and extinction. As the number of interpolated trials becomes large, resistance to extinction should approach that of continuously reinforced *Ss*. The present experiment was designed to test these predictions from the discrimination theory of partial reinforcement.

METHOD

Apparatus.—The apparatus was a black wooden runway 5 in. wide, 5 in. high, and 99 in. long. The first 12 in. comprised the start box and the last 15 in. the goal box. The start and goal boxes were separated from the alley by metal guillotine doors. A 1.8 × 2.5 in. food cup was concealed behind a 1.6-in. metal shield at the end of the goal box. The presence of the shield required *S* to put his head into the food cup to see its contents.

Response time was measured in .01 sec. from the opening of the start door until *S* interrupted a photobeam 5 in. inside the goal box. On rewarded trials a food pellet was automatically delivered by a solenoid feeder when *S* interrupted a second photobeam positioned across the food cup. The reward consisted of one "Frostyos" pellet, a sugar-coated breakfast cereal produced by General Mills, Incorporated. On nonrewarded trials the solenoid operated, but no food was delivered. The *Ss* required about 15 sec. to consume the pellet, and were removed to the home cage 15 sec. after the interruption of the first photobeam.

Subjects.—The *Ss* were 60 naive Slonaker albino rats, 28 males and 32 females, selected from the colony maintained by the Stanford University Psychology Department. They were approximately 90 days old at the beginning of the experiment and had been tamed for 3 days prior to the experiment. The *Ss* were housed in individual cages with ad lib. water and were maintained on 22 hr. food deprivation throughout the experiment. The *Ss* were permitted to eat lab chow for 2 hr. after each daily experimental session.

Procedure.—After taming, *S* was placed into the goal box for 5 trials. Reward pellets were scattered throughout the goal box on the first 2 trials, but only the food cup contained pellets on the last 3 trials. When *S* ate one pellet he was removed to his home cage. Preliminary runway training followed on the next 6 days at 5 trials a day. The 30 preliminary runway trials were continuously rewarded for all *Ss*. The *Ss* were run in rotation, and the intertrial interval decreased from approximately 60 to 30 min. during this preliminary training.

After preliminary training, *Ss* were divided into five groups, equated on performance during the preliminary 30 trials. During the remaining part of the experiment *Ss* received 5 runway trials each day. Three groups of 14 *Ss* each received 40% partial reinforcedment (P) for 70 trials. The schedule was random except for the restrictions that runs of longer than 6 rewarded or nonrewarded trials were excluded and that the series ended with a rewarded trial. One of the P groups was reduced to 13 *Ss* when a rat died during the experiment. The remaining two groups, each consisting of 9 *Ss*, received continuous reinforcement (C) for 70 trials. All *Ss* were left in the goal box for 15 sec. on each trial during this phase, irrespective of whether the trial was rewarded or not. The intertrial intervals averaged about 10 min.

Groups P-O and C-O were extinguished following the last of the 70 training trials. Group P-25, however, received 25 additional 100% rewarded trials, and then was extinguished. Groups P-70 and C-70 were run for 14 additional days on 100% reward, receiving 70 additional continuously rewarded trials before they were extinguished.

During extinction *S* received 5 trials each day until he accumulated 5 trials, not necessarily consecutive, during which he did not enter the goal box within 60 sec. In order to

obtain extinction curves based on all *S*s, each *S* was run for at least 40 trials. If *S* did not enter the goal box within 60 sec., he was removed to his home cage and given a score of 60 sec. If *S* entered the goal box, he was detained in the box for 15 sec. The empty feeder operated if *S* looked into the food cup. The intertrial interval averaged about 10 min.

RESULTS

Training.—Curves of response speed during training are presented in Fig. 1. An *S*'s median response speed for each block of 5 trials was computed. Each point on the curves represents a group mean of these median scores for a given block of trials. The *S*s were divided into five groups just before Block 7. An analysis of variance on the median speed scores of Block 7 yielded an $F < 1.0$. Hence, the division of *S*s did result in comparable groups.

The curves show that between Blocks 7 and 20, the trials on which the P *S*s received 40% reinforcement, the continuously reinforced *S*s eventually ran faster than the partially reinforced *S*s. The means and *SD*s of *S*s' median speeds during Block 20 are presented in Table 1. An analysis of variance performed on these scores indicated that there was no significant difference between the two continuously reinforced groups ($F = .28$, $df = 1/54$), nor among the three partial groups ($F = 1.14$, $df = 2/54$). However, the partial groups ran significantly slower than the 100% groups ($F = 7.16$, $df = 1/54$, $P < .01$).

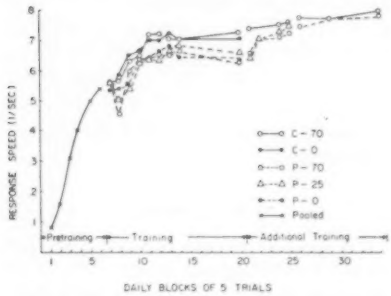

FIG. 1. Response speeds during training.

TABLE 1

MEANS AND *SD*s OF SPEED SCORES ON BLOCK 20

Group	Mean	*SD*	*N*
P-0	.64	.07	13
P-25	.66	.07	14
P-70	.62	.08	14
C-0	.70	.09	9
C-70	.72	.11	9

After Block 20, when Groups P-25 and P-70 were changed from 40% reinforcement to 100%, their response speeds increased essentially to that of the C-70 *S*s. Since Group P-25 was extinguished following Block 25, a comparison of the P groups and Group C-70 was made by an analysis of variance on the Block 25 median speed scores. There was no significant difference between the two P groups ($F = 1.13$, $df = 1/34$) nor between the combined P groups and Group C-70 ($F = 1.00$, $df = 1/34$). Thus, while on 40% reinforcement, the P *S*s ran significantly slower than the 100% *S*s. When the P *S*s were subsequently given 100% reinforcement, however, their response speeds significantly increased so that the initial difference between the 40% and 100% *S*s disappeared in less than 25 trials.

Extinction.—Curves of response time during extinction are given in Fig. 2. Each point on the curves represents a group mean of individual median time scores for a given block of 5 extinction trials. The striking feature of Fig. 2 is the difference between the extinction curves of the P and C groups. The three P groups, all of which initially received 40% reinforcement, have very similar extinc-

tion curves in spite of the fact that one group had 0, the second group 25, and the third 70 trials of 100% reinforcement between partial reinforcement and extinction. The curves of the two C groups are also very similar to each other. The difference between the P and C groups is evident as early as the second block of 5 extinction trials and indicates much faster extinction for the C groups. Thus, additional trials of 100% reinforcement following 40% reinforcement had little effect on rate of extinction.

The large difference in extinction rate between the P and C groups evidenced by the curves is also reflected in the mean number of trials to reach the extinction criterion (Table 2). All three P groups took at least 50 trials to reach the criterion, but the C groups took only about 32 trials. It can be seen that the *SD*s for the C groups are much smaller than those for the P groups, making inadvisable the use of an overall analysis of variance on the responses to the criterion scores. For Group C-0 and Group C-70, a t of .78 was obtained, indicating that, for 100% reinforced *S*s, a difference between 100 and 170 training trials had no effect on resistance to extinction. Homogeneity of variance among the responses to the criterion for the three P groups permitted a 1×3 analysis of variance. The obtained F of 6.26 ($df = 2/38$) is significant at the .01 level, and indicates that individual t tests are appropriate. Using the overall within-groups variance as the variance estimate for the error term of the t tests, Groups P-0 and P-25 are not significantly different from each other ($t = .88$). The t between Groups P-0 and P-70 of 2.46, however, is significant ($df = 25$, $P < .05$). Similarly, the t between Groups P-25 and P-70 of 3.40 is significant ($df = 26$, $P < .01$).

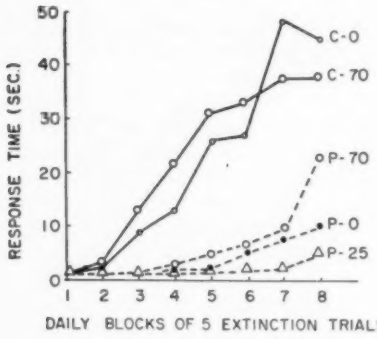

FIG. 2. Response times during extinction.

TABLE 2

MEANS AND *SD*s OF RESPONSES TO THE EXTINCTION CRITERION

Group	Mean	*SD*	*N*
P-0	61	14	13
P-25	66	11	14
P-70	50	10	14
C-0	33	5	9
C-70	31	6	9

In order to demonstrate that all the P groups are significantly more resistant to extinction than the C groups, a median test was run between Group P-70, the least resistant of the P groups, and the combination of the two C groups. The obtained χ^2 of 24.89 is highly significant ($P < .001$, $df = 1$). There is almost no overlap in the two distributions. This large difference was obtained in spite of the fact that Group P-70 had 70 continuously rewarded trials just prior to extinction.

DISCUSSION

The present results indicate that the partial reinforcement effect (PRE) of increased resistance to extinction is relatively unaffected by continuous rein-

forcement following the partial reinforcement. After the present experiment was completed, a study by Quartermain and Vaughan (1961) was published which bears on the same issue. Using lever pressing in a Skinner box, they compared resistance to extinction of rats given acquisition training with either 800 responses of which a random 10% were reinforced, or 400 responses at 10% reinforcement followed by 400 continuously reinforced responses. Their results were entirely consistent with those reported here. These findings demonstrate that discriminability of the transition from training to extinction cannot be taken as explaining much of the partial reinforcement effect.

Although it could be assumed that discrimination has some small effect since in the present study Group P-70 extinguished slightly faster than Group P-0, the fact that there was no difference in resistance to extinction between Groups P-0 and P-25 presents a problem for the discrimination hypothesis. It might be argued that 25 trials were not sufficient for *S*s to have noticed a change in the reinforcement schedule, and, hence, a difference between Groups P-0 and P-25 should not have been expected. However, the schedule change clearly influenced running performance. The response speeds of Group P-25 and P-70 *S*s increased markedly during the 25 continuously reinforced trials, eventually equaling the speeds of the Group C-70 *S*s. This increase in performance can be taken as evidence that the *S*s "noticed" the change.

Another proposal might be that Group P-25's resistance to extinction could have been increased by the fact that it had 25 more trials than Group P-0. However, Group C-70 had 70 more trials than did Group C-0, and these C groups did not differ in resistance to extinction. Thus, with the degree of training given in the present experiment, additional trials had no effect on resistance to extinction.

The present experiment can be added to a growing amount of empirical data indicating that the discrimination hypothesis cannot adequately account for the PRE. Marx (1958) has shown that if certain significant features of the goal situation (i.e., the food cup) are omitted on half of the extinction trials, rate of extinction is decreased rather than increased as would be predicted from a discrimination analysis. The discrimination theory has been most successful in accounting for the fact that rate of extinction usually varies directly with the percentage of reward when number of trials or number of reinforcements are held constant. However, using a design which pitted percentage of reinforcement against number of nonreinforcements, Lawrence, Festinger, and Theios (summarized by Festinger, 1961) found that resistance to extinction was independent of the reinforcement percentage, but was determined primarily by the *number* of nonreinforced trials *S*s experienced during training.

The present data as well as these other studies suggest that the discrimination hypothesis is inadequate as an interpretation of partial reinforcement. The present data indicate that an adequate theory of partial reinforcement must have constructs representing relatively permanent effects of nonreinforcement which can be sustained through blocks of continuous reinforcement. At least four existing theories of partial reinforcement employ constructs of this character. These theories include Amsel (1958), Estes (1959), Logan (1960), and Festinger (1961). The present results do not enable one to discriminate among these theories.

Summary

Three groups of 14 rats each received 70 trials of random 40% reinforcement in a runway. Then different numbers of continuously reinforced trials were interpolated between the 40% reinforcement and extinction. Group P-0 received none, Group P-25 received 25, and Group P-70 received 70 interpolated 100% reinforced trials. Two control groups ($N = 9$ each) received only continuous rein-

forcement during training: Group C-0 was extinguished with partial Group P-0; Group C-70 was given 70 additional reinforced trials and extinguished with Group P-70.

There was no significant difference in resistance to extinction between the two continuously reinforced groups. Group P-70 was very significantly ($P < .001$) more resistant to extinction than the two continuously reinforced groups. There was no significant difference in resistance to extinction between Groups P-0 and P-25, but both of these groups were slightly more resistant than Group P-70. These results question the adequacy of the discrimination hypothesis that the partial reinforcement extinction effect results from difficulty in discriminating the transition from training to extinction.

References

Amsel, A. The role of frustrative nonreward in noncontinuous reward situation. *Psychol. Bull.*, 1958, **55**, 102–119.

Bitterman, M. E., Fedderson, W. E., & Tyler, D. W. Secondary reinforcement and the discrimination hypothesis. *Amer. J. Psychol.*, 1953, **66**, 456–464.

Estes, W. K. The statistical approach to learning theory. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. New York: McGraw-Hill, 1959.

Festinger, L. The psychological effects of insufficient reward. *Amer. Psychologist*, 1961, **16**, 1–11.

Logan, F. A. *Incentive.* New Haven: Yale Univer. Press, 1960.

Marx, M. H. Resistance to extinction as a function of continuous or intermittent presentation of a training cue. *J. exp. Psychol.*, 1958, **56**, 251–255.

Mowrer, O. H., & Jones, H. M. Habit strength as a function of the pattern of reinforcement. *J. exp. Psychol.*, 1945, **43**, 293–311.

Quartermain, D., & Vaughan, G. M. Effect of interpolating continuous reinforcement between partial training and extinction. *Psychol. Rep.*, 1961, **8**, 235–237.

(Received June 6, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 1, 7–19

THE EFFECTS OF LANGUAGE HABITS ON THE ACQUISITION AND RETENTION OF VERBAL ASSOCIATIONS [1]

LEO POSTMAN

University of California

The experiments reported in this paper investigate the effects of unit-sequence habits on the acquisition and retention of verbal materials. By unit-sequence habits we refer to associative connections between verbal items established through linguistic usage (Underwood & Postman, 1960). The transfer effects of such language habits may be either positive or negative, depending on the relationship between *S*'s pre-experimental associations and those prescribed in the learning task.

The amount of transfer from pre-experimental habits should increase with the frequency of usage of the verbal units in the language. The higher the frequency of usage, the larger on the average is the number of different contexts in which an item is likely to appear and hence the larger the number of different associations which it acquires. Conventional indices of meaningfulness such as association value and *m* value do in fact correlate highly with frequency of usage (Postman, 1961; Underwood & Schulz, 1960). Both unit-sequence facilitation and unit-sequence interference may then be expected to increase as a function of frequency of usage. When the sequence to be learned involves high-frequency items, at least some of the prescribed associations are likely to agree with pre-experimental habits or can be readily established through direct mediational links. However, the number of pre-experimental associations which can compete with the prescribed ones will also increase as a function of frequency of usage. With low-frequency materials, the prescribed sequences will have little pre-experimental strength and few direct mediators will be available, but there will also be few pre-experimental associations to produce interference.

The conditions of facilitation and interference should influence recall as well as acquisition. When there is unit-sequence facilitation, it is assumed that the strength of the pre-experimental habits has generalized to the prescribed associations within the list. The greater the generalized habit strength, the more resistant the items in the experimental list will be to interference. When there is unit-sequence interference, it is assumed that competing pre-experimental habits have been unlearned during acquisition and have recovered over time to become sources of interference at recall.

An earlier study (Postman, 1961) investigated the acquisition and retention of serial lists of words of high and low frequency of usage (Lists HF and LF). List HF was learned faster than List LF. At the same time, however, the rate of misplaced responses was substantially greater for List HF. The positive relationship between speed of learning and error rate points to covariation of unit-sequence facilitation and interference as a function of word frequency. There were no significant differences in the amount of retention for the two kinds of materials. However, misplaced responses at recall recovered at a faster rate for List HF than for List LF. While there was no evidence for superior retention of materials of high meaningfulness, the original expectation of an inverse relationship between word frequency and recall was not borne out. This expectation had been based on the

[1] This research was supported by a grant from the National Science Foundation.

assumption that competition from extra-experimental associations would increase directly with word frequency. The difference in recovery of errors is in accord with this assumption. However, the fact that the amounts recalled were comparable in spite of a substantial difference in the incidence of overt errors suggests that increases in unit-sequence interference may be offset in part by unit-sequence facilitation.

The present experiments extend the analysis of the conditions of unit-sequence facilitation and interference. A first step was to vary the word frequency of stimuli (Ss) and responses (Rs) in a study investigating the acquisition and retention of paired associates (Exp. I). There is considerable evidence that speed of learning varies directly with the meaningfulness of Ss and Rs, and that this relationship is considerably more pronounced and consistent for Rs than for Ss (Cieutat, Stockwell, & Noble, 1958; Hunt, 1959; Kimble & Dufort, 1955; L'Abate, 1959; Mandler & Campbell, 1957; Morikawa, 1959; Underwood & Schulz, 1960). In these experiments the units scaled for meaningfulness (e.g., trigrams or paralogs) were either nonsense items or, more usually, included both nonsense items and words. Letter-sequence habits as well as unit-sequence habits contribute to the differences in speed of learning obtained with such materials (Underwood & Schulz, 1960). The lists used in Exp. I consisted entirely of words so that variations in speed of learning and amount of retention could be attributed primarily to the influence of unit-sequence habits. It was expected that both unit-sequence facilitation and unit-sequence interference would increase with word frequency, with the relative amounts of interference growing more rapidly as a function of S frequency than of R frequency. To the extent that paired items are practiced in the order of their presentation, competing associations elicited by the Ss should be more effective in delaying acquisition than those evoked by the Rs. These considerations suggest that speed of learning should rise to a maximum and then decline as the word frequency of Ss is increased. On the other hand, speed of learning should vary directly with the word frequency of Rs.

No prior studies of the effects of S and R meaningfulness on the long-term retention of paired associates are available. If the associative context of the Ss rather than the Rs is critical in determining the amount of effective competition at recall, differences in the amount of retention should be determined largely by the degrees of unit-sequence facilitation and interference falling on the S terms. Thus, the differences among S conditions in learning should predict recall better than those among the R conditions.

According to the present analysis, a sample of high-frequency Ss will activate strong pre-experimental habits whereas a sample of low-frequency Ss will not. It follows that manipulation of the conditions of facilitation and interference within a list should be more effective with low-frequency Ss than with high-frequency Ss. When pre-experimental associative probabilities are low, the introduction of otherwise unlikely sources of facilitation and interference based on language habits should produce major effects on learning and retention. On the other hand, when the associative probabilities are high, the introduction of known associates or competitors should add relatively little to the amounts of facilitation and interference from uncontrolled sources. This prediction was tested in Exp. II in which the pre-experimental probability of the prescribed associations was varied for Ss of high and low word frequency.

Experiment III was designed to provide a further test of the assumption that the higher the word frequency of a series of items, the greater is the amount of overlap among their associative contexts. The method of verbal discrimination learning was used. To the extent that discriminability depends on the availability of differential responses to the members of a pair, speed of discrimination learning should increase with the meaningfulness of the paired items.

TABLE 1

STIMULUS AND RESPONSE TERMS IN EXP. I

S_h	R_h	S_m	R_m	S_l	R_l
BUILDING	ANSWER	ARBOR	BASIN	BRAMBLE	ABBESS
COUNTRY	COLOR	BURLAP	DOGMA	CAUCUS	BUFFOON
DOCTOR	DINNER	CINDER	FETISH	DECOY	DOTAGE
GARDEN	FIGURE	DISCORD	HERMIT	FARTHING	HAREM
HUSBAND	MOMENT	GROCER	MINSTREL	GULLET	OBOE
LETTER	PROBLEM	LOTION	OMEN	LORRY	PREFIX
MORNING	REASON	MAGNATE	RELIC	OXIDE	RAMROD
PAPER	SHOULDER	OATMEAL	SUFFRAGE	PESTLE	STANZA
STORY	TABLE	TRAITOR	TEMPEST	SEQUEL	TENURE
WINDOW	WOMAN	WHISKER	WAFER	WICKET	WAMPUM

Such a relationship was found with pairs of nonsense syllables by Runquist and Freeman (1960). However, as the associates elicited by the items increase in number, they should begin to lose distinctiveness because of the increasing probability of overlap in associative context. Acquisition will then require the unlearning of differential responses which produce errors of generalization, but such responses may be expected to recover with the passage of time. Thus, speed of acquisition and amount of retention for the verbal discrimination task should first increase and then decrease as a function of word frequency.

EXPERIMENT I

Method

Materials.—The learning materials were two-syllable nouns of high, medium, and low frequency of usage. Each frequency range was represented by two lists of 10 words each, one of which served as the S list and the other as the R list (S_h, S_m, S_l; R_h, R_m, R_l). These materials were used in the construction of nine paired-associate lists representing all the possible combinations of word frequencies of the S and R items. The lists of words are shown in Table 1.

The lists sample three frequency ranges in the "L" count of Thorndike and Lorge (1944). The numbers of occurrences in 4.5 million are 1000–3300, 10–33, and 1–3 for words of high, medium, and low frequency, respectively. The mean rated familiarity and *m* values, which were obtained in prior standardization procedures, are correlated with frequency of usage. The words were chosen so as to minimize pre-experimental associations between items in the S lists and items in the R lists. Norms of free association obtained from 1000 students at the University of California were used for this purpose.

There was no duplication of first letters in any of the S or R lists. For all nine combinations of word frequencies the S list and the R list had six first letters in common. Four different pairings of Ss and Rs were used equally often with each of the nine lists.

Procedure.—The lists were presented on a Hull-type memory drum at a 2:2 rate (S alone for 2 sec. and S and R together for 2 sec.), with an 8-sec. intertrial interval. Learning was to a criterion of one perfect trial. Four different orders of presentation were used to minimize serial learning. Each of the four random orders was used as a starting order equally often.

Retention was tested by relearning either 30 sec. or 7 days after the end of original practice. Relearning was for five trials or to criterion, whichever took the longer.

Subjects.—With nine types of lists and two retention intervals, there were 18 groups of 16 *Ss* each. All *Ss* were undergraduate students who had English as their native language and were naive to rote-learning experiments. The *Ss* were assigned to conditions in blocks of 18, with 1 *S* per block for each combination of lists and retention intervals. The running order within each block was determined by a table of random numbers, as was the assignment to different pairings and starting orders. No *Ss* were lost because of failure to learn.

Results

Original learning.—There were no significant differences between the

TABLE 2

MEAN NUMBERS OF TRIALS TO CRITERION: EXP. I

R Terms	Stimulus Terms						Over-All Mean
	S_h		S_m		S_l		
	Mean	*SD*	Mean	*SD*	Mean	*SD*	
R_h	10.5	6.3	10.5	40.	12.1	6.9	11.0
R_m	12.9	6.0	12.2	54.	15.7	8.8	13.6
R_l	16.2	8.5	11.2	4.0	16.3	6.5	14.6
Over-all Mean	13.2		11.3		14.7		

30-sec. and 7-day groups learning a given type of list. The results of the two groups learning the same materials have, therefore, been combined. Table 2 shows the mean trials to criterion for each of the nine kinds of lists and also summarizes the average trends produced by the word frequency of the Ss and Rs. Speed of learning first increases and then decreases as a function of the word frequency of Ss but varies directly with the word frequency of Rs. The over-all effects of the word frequency of both Ss and Rs are significant beyond the .01 level ($F = 6.15$ for Ss and 7.24 for Rs, $df = 2/279$ for both). The S × R interaction is not significant ($F = 1.54$, $df = 4/279$).

Errors during learning.—Table 3 shows the mean percentages of misplaced Rs during learning. Such responses account for the great majority of overt errors. The percentages are based on opportunities (total number of presentations minus number of correct responses) and thus are independent of speed of learning.

For both Ss and Rs the mean percentages of errors are clearly larger when word frequency is high than when it is medium or low. There are only small differences between the latter two conditions. Following arc-sine transformation and with 2/279 *df*, the *F* ratio for Ss is 12.44 ($P < .001$) and 3.11 for Rs ($.02 < P < .05$). There is also a significant S × R interaction ($F = 2.62$, $df = 4/279$, $.02 < P < .05$). When the word frequency of either the Ss or the Rs is high, the percentage of errors remains uniformly large, regardless of the value of the other variable.

Table 3 also shows for each type of list the product-moment correlation between trials to criterion and percentage of misplaced Rs during learn-

TABLE 3

MEAN PERCENTAGES OF MISPLACED RESPONSES AND CORRELATION BETWEEN PERCENTAGES AND TRIALS TO CRITERION: EXP. I

R Terms	Stimulus Terms									Over-All Mean
	S_h			S_m			S_l			
	Mean	*SD*	*r*	Mean	*SD*	*r*	Mean	*SD*	*r*	
R_h	26.6	17.5	.61*	23.0	17.9	.55*	26.3	17.6	.55*	25.3
R_m	27.1	12.9	.26	16.4	10.1	.59*	16.9	12.1	.30	20.1
R_l	30.0	15.3	.34	14.8	12.1	.61*	14.7	7.6	.26	19.8
Over-all Mean	27.9			18.1			19.3			

* $P < .01$.

TABLE 4

MEAN NUMBERS OF ITEMS RECALLED IN EXP. I

R Terms	30-Sec. Recall							7-Day Recall						
	Stimulus Terms						Over-All Mean	Stimulus Terms						Over-All Mean
	S_h		S_m		S_l			S_h		S_m		S_l		
	Mean	*SD*	Mean	*SD*	Mean	*SD*		Mean	*SD*	Mean	*SD*	Mean	*SD*	
R_h	9.4	.7	9.2	.7	9.2	.8	9.3	5.3	1.7	5.4	2.1	5.2	1.9	5.3
R_m	9.2	.7	8.8	.8	9.0	1.2	9.0	4.9	2.0	5.1	1.9	3.8	2.1	4.6
R_l	8.7	.9	9.1	1.2	8.6	1.0	8.8	4.7	2.3	5.2	1.9	3.5	2.4	4.5
Over-all Mean	9.1		9.0		8.9			5.0		5.2		4.2		

ing. All the correlations are positive, i.e., the higher the error rate, the more slowly is criterion reached. This relationship is especially pronounced and statistically significant for Cond. S_m and R_h.

The inverse relationship between error rate and speed of learning found for individuals within groups clearly does not hold for materials. On the R side, both speed of learning and error rate are maximal for Cond. R_h; on the S side, error rate is maximal, and speed of learning is intermediate, for Cond. S_h. Within a given group, however, rate of acquisition varies with *Ss*' susceptibility to interference, especially when learning is fast as in Cond. S_m and R_h. When correct Rs are acquired rapidly, a low percentage of misplaced Rs is favorable to attainment of the criterion, i.e., failures to respond are eliminated more readily than overt errors. When correct Rs increase relatively slowly, the criterion score is less sensitive to the type of error.

Recall.—Table 4 shows the mean numbers of Rs recalled by the various groups. The drops on the 30-sec. test vary inversely with speed of learning although the correlation is not perfect. The amounts of retention loss (mean 30-sec. recall − mean 7-day recall) as a function of the word frequency of Ss and Rs are shown in Fig. 1. The losses during the 7-day interval are smallest for S_m and greatest for S_l, with S_h yielding an intermediate amount of forgetting. The

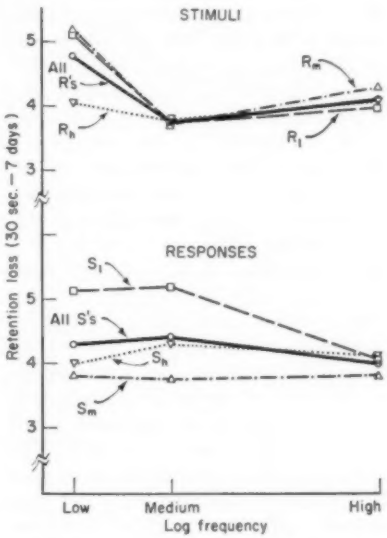

FIG. 1. Amounts of retention loss (mean 30-sec. recall − mean 7-day recall) as a function of the word frequency of Ss and Rs (Exp. I).

TABLE 5

NUMBERS OF MISPLACED RESPONSES AT RECALL: EXP. I

R Terms	30-Sec. Recall								7-Day Recall							
	Stimulus Terms						Over-All		Stimulus Terms						Over-All	
	S_h		S_m		S_l				S_h		S_m		S_l			
	Total	*N*	Total	*N*	Total	*N*	Total	*N*	Total	*N*	Total	*N*	Total	*N*	Total	*N*
R_h	5	4	8	7	5	5	18	16	17	8	12	7	13	9	42	24
R_m	5	5	9	7	4	4	18	16	9	6	4	4	9	8	22	18
R_l	11	9	2	1	6	5	19	15	9	6	4	2	10	6	23	14
Over-all	21	18	19	15	15	14			35	20	20	13	32	23		

Note.—*N* = Number of *Ss* giving misplaced responses.

average difference between S_h and S_m is, however, clearly smaller than that between S_l and S_m. The trends for the three values of R are comparable. There are no consistent variations in retention loss as a function of the word frequency of Rs. The slight over-all advantage of R_h is due entirely to the variations under Cond. S_l and is not present at the other two values of S. Following a Freeman-Tukey square root transformation, the retention scores were subjected to an analysis of variance. The significance of the differences in the amount of forgetting is assessed by the interactions of time (T) with the values of S and R. Only the interaction, S × T, is significant ($F = 3.08$, $df = 2/270$, $.02 < P < .05$). The F ratios for R×T and S×R×T are .88 and 1.08, respectively. Thus, the rank order of the S conditions is the same for speed of original learning and amount of long-term retention. On the other hand, the differences in rate of acquisition among the R conditions are not reflected consistently in the amounts of retention.

Misplaced responses at recall.—The numbers of misplaced Rs at recall and the numbers of *Ss* contributing them are shown in Table 5. Such errors account for only a small fraction of the total retention losses. A large proportion of *Ss*—99/144 in the 30-sec. groups and 88/144 in the 7-day groups—failed to give misplaced Rs. The observed frequencies indicate a higher rate of recovery of intralist intrusions for S_h and S_l than for S_m, and also a higher rate for R_h than for R_m and R_l. In view of the high proportion of zero scores, the differences in temporal trends were not evaluated statistically and must be interpreted with caution. It is interesting to note, however, that the apparent differences in rate of recovery are related to the amount of forgetting for the S conditions but not the R conditions. The low rate of intralist intrusions suggests that *Ss* had considerable success in rejecting incorrect Rs.

Relearning.—The mean numbers of trials to criterion in relearning are shown in Table 6. The relative amounts of retention loss correspond to those found at recall: the increase in trials to criterion is greater for S_h and S_l than for S_m whereas the word frequency of Rs has little effect. The interaction, S × T, is significant ($F = 4.52$, $df = 2/270$, $.02 < P < .05$).

TABLE 6

MEAN NUMBERS OF TRIALS IN RELARNING TO CRITERION: EXP. I

R Terms	30-Sec. Recall							7-Day Recall						
	Stimulus Terms						Over-All Mean	Stimulus Terms						Over-All Mean
	S_h		S_m		S_l			S_h		S_m		S_l		
	Mean	*SD*	Mean	*SD*	Mean	*SD*		Mean	*SD*	Mean	*SD*	Mean	*SD*	
R_h	2.2	1.7	2.6	2.2	2.1	1.2	2.3	5.1	1.8	3.8	1.8	4.1	1.9	4.3
R_m	2.1	.9	3.1	2.2	2.4	2.1	2.5	4.4	2.5	3.5	1.4	5.9	2.8	4.6
R_l	2.5	1.1	1.9	1.0	2.8	1.4	2.4	5.2	2.5	3.2	1.2	5.1	4.3	4.5
Over-all Mean	2.3		2.5		2.4			4.9		3.5		5.0		

The *F* ratios for R×T and S×R×T are .01 and .85, respectively.

EXPERIMENT II

Method

Materials.—There were two lists of 10 paired associates, one with S terms of high word frequency (List S_h), and the other with S terms of low word frequency (List S_l). The S terms in both lists were two-syllable nouns drawn from the same pools of items as the words in Exp. I. Half the Rs in each list were highly probable associates (A_h) of the S words, i.e., they were the primary Rs to these words in free association norms obtained from 1000 students at the University of California. The other half of the Rs were associates of low probability (A_l), having been given by only 2/1000 *Ss* in the normative group. Thus, there were four types of pairs: S_h-A_h and S_h-A_l in one list, S_l-A_h and S_l-A_l in the other. All the Rs were nouns of high frequency of usage falling within the AA and A ranges of the General Count of Thorndike and Lorge (1944). Within each of the four classes of pairs three of the Rs were two-syllable words and two Rs were one-syllable words.

There was no duplication of first letters in either of the S lists. One first letter was duplicated in each of the two R lists. In both

TABLE 7

STIMULUS-RESPONSE PAIRS USED IN EXP. II

Stimulus	A_h	Assoc. Prob.*	Stimulus	A_l	Assoc. Prob.*
List S_h					
GARDEN	FLOWER	592	BUILDING	SIZE	2
MOMENT	TIME	255	COUNTRY	VALLEY	2
ORDER	COMMAND	115	LETTER	PERSON	2
TROUBLE	PROBLEM	132	REASON	NEED	2
WINDOW	GLASS	252	SHOULDER	BURDEN	2
List S_l					
CAUCUS	MEETING	156	BRAMBLE	PAIN	2
GULLET	THROAT	180	DECOY	ARMY	2
HAREM	WOMAN	340	OXIDE	COLOR	2
RAMROD	GUN	258	TENURE	LIFE	2
STANZA	MUSIC	282	WAMPUM	DOLLAR	2

* Number of *Ss* out of 1000 giving response on test of free association.

TABLE 8

MEAN NUMBERS OF TRIALS TO CRITERION OF 5/5 FOR EACH CLASS OF PAIRS AND 10/10 FOR TOTAL LISTS IN ORIGINAL LEARNING AND RELEARNING: EXP. II

Pairs	30-Sec. Groups Stimulus List				7-Day Groups Stimulus List			
	S_h		S_l		S_h		S_l	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
Original Learning								
A_h	4.6	2.7	1.9	.9	5.4	3.7	1.6	.5
A_l	4.3	2.0	2.8	1.0	4.2	1.5	2.6	1.5
Total list	5.4	2.9	2.9	1.0	6.2	3.4	2.8	1.5
Relearning								
A_h	1.4	.8	1.0	0	3.3	1.9	1.7	.6
A_l	1.8	1.0	1.2	.6	2.6	1.3	2.2	.6
Total list	1.9	1.2	1.2	.6	4.7	1.9	2.5	.8

lists the S and R terms had six first letters in common. Within a list none of the Rs appeared in the association tables for any of the stimuli except the one with which it was paired. Table 7 presents the lists and for each pair shows the frequency of the R term in the association norms.

Procedure.—The lists were presented on a Hull-type memory drum at a 2:2 rate with an 8-sec. intertrial interval. Learning was to one perfect trial. There were four different orders of presentation each of which was used as a starting order equally often. Retention was tested by relearning either 30 sec. or 7 days after the end of original learning. Relearning was for five trials or to criterion, whichever took the longer.

Subjects.—With two lists and two retention intervals, the design comprised four groups of 16 *Ss* each. All *Ss* were undergraduate students who had English as their native language and were naive to rote-learning experiments. The *Ss* were assigned to conditions in blocks of 4, with 1 *S* per block for each combination of lists and retention intervals. Assignment to conditions within a block was determined by a table of random numbers as was the assignment to different starting orders. No *Ss* were lost because of failure to learn.

Results

Original learning.—The mean numbers of trials to a criterion of 5/5 for each type of pair as well as to a criterion of 10/10 for the entire list are shown in Table 8. There were no significant differences between the 30-sec. and 7-day groups learning a given kind of list. Pairs S_l-A_h were mastered extremely rapidly, with the large majority of *Ss* reaching the criterion of 5/5 in either one or two trials. Pairs S_l-A_l were learned more slowly than Pairs S_l-A_h. However, with half the pairs learned almost immediately, acquisition of the entire list was very rapid. By contrast, Pairs S_h-A_h were not learned faster than Pairs S_h-A_l; in fact, there was a trend in the opposite direction. The over-all rate of learning was significantly slower for List S_h than for List S_l ($t = 4.97$, $df = 62$, $P < .01$). Analysis of variance of the criterion scores on the four kinds of pairs shows the interaction, Pairs × Lists, to be significant ($F = 7.08$, $df = 1/62$, $P < .01$). Thus, the association norms predict speed of learning for List S_l but not for List S_h.

Errors during learning.—Misplaced Rs were much more common during

the acquisition of List S_h than of List S_l. Twenty-two of the *Ss* learning List S_h made such errors, and only 6 of the *Ss* learning List S_l. When the numbers of misplaced Rs are expressed as percentages based on opportunities, the mean percentage for List S_h is 27.6 (SD = 20.0; Median = 26.8). For List S_l the mean of the extremely skewed distribution is 6.4 (SD = 16.1; Median < 1.0). The difference between the two distributions is significant beyond the .01 level by Wilcoxon's test for signed ranks, with *Ss* in the same block treated as paired replicates.

A total of 23 importations from outside the list was contributed by 13 of the *Ss* learning List S_l; 3 of the *Ss* learning list S_h contributed a total of five such errors. The difference between the proportions of *Ss* making such errors is significant beyond the .01 level ($\chi^2 = 6.75$; $df = 1$). All of the importations appeared in the association tables for the appropriate stimuli. Of 23 importations into List S_l, 19 were substitutions of primary associates for prescribed A_l responses; no such cases occurred during the learning of List S_h.

Recall.—The recall scores for the four kinds of pairs are shown in Table 9. When we consider the two types of pairs within each list, we find clearly superior retention of the high-probability pairs in List S_l but no difference in List S_h. Comparison of the corresponding pairs in the two lists shows retention to be higher for Pairs S_l-A_h than for Pairs S_h-A_h. By contrast retention is poorer for Pairs S_l-A_l than for Pairs S_h-A_l.

In view of the virtual absence of any variation in the recall scores on the 30-sec. test, the following procedure was used to evaluate the statistical significance of the differences in the amount of forgetting. Within each block, the difference between the scores obtained by *Ss* in the 30-sec. and 7-day groups was determined for each type of pair, and the distribution of differences was subjected to an analysis of variance. The over-all difference between Lists is not significant ($F < 1$), whereas the interaction of Lists with Pairs is significant beyond the .01 level ($F = 8.59$, $df = 1/30$). The relationship between the associative rank of the Rs and the rate of forgetting varies reliably with the word frequency of the Ss.

Errors at recall.—The numbers of misplaced Rs at recall were small and did not increase during the retention interval. The few intralist errors that were made occurred almost exclusively during the recall of List S_h, at the rate of approximately .5 per *S* on both tests. By contrast, importations from outside the list showed a systematic temporal trend. On the 30-sec. test there were only 2 *Ss* giving such Rs, both on Pairs S_l-A_l. The frequencies of importations on the 7-day test and the numbers of *Ss* (N) contributing them were as follows: S_h-A_h—4 (N = 4); S_h-A_l—5 (N = 5); S_l-A_h—5 (N = 5); S_l-A_l—19 (N = 13). When the 30-sec. and 7-day *Ss* assigned to the same block are compared, the number of cases in which there is an increase in

TABLE 9

MEAN NUMBERS OF PAIRS RECALLED IN EXP II

Pairs	30-Sec. Recall Stimulus List				7-Day Recall Stimulus List			
	S_h		S_l		S_h		S_l	
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
A_h	4.8	.4	5.0	0	3.2	1.2	4.1	.8
A_l	4.5	.5	4.8	.4	3.0	1.3	2.2	1.2

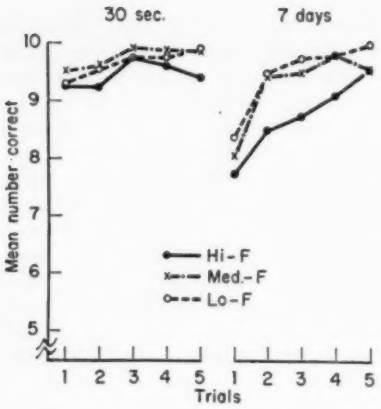

FIG. 2. Mean numbers of correct responses in relearning of verbal discrimination (Exp. III).

importations during the retention interval is significantly greater for List S_l than for List S_h ($\chi^2 = 4.52$, $df = 1$, $.02 < P < .05$). It is clear that this difference is primarily a function of the number of imported substitutions for low-probability associates. In the recall of List S_h, none of these substitutions were primary associates of the appropriate Ss; in the case of List S_l, 16/19 substitutions were primary associates.

Relearning.—The mean numbers of trials to relearn to a criterion of 5/5 on each type of pair and 10/10 for the entire list are shown in Table 8. The differences in the amount of forgetting as measured by speed of relearning correspond to those found at recall. The procedure used in analyzing the temporal trends in recall was also applied to the relearning scores. The over-all difference between Lists was again not reliable ($F = 1.53$), whereas the interaction of Lists with Pairs is significant ($F = 5.11$, $df = 1/30$, $.02 < P < .05$).

EXPERIMENT III

Method

Materials.—The learning materials were three lists of 10 pairs of words each. One list (List HF) was composed of words of high frequency of usage, one list (List MF) of words of medium frequency, and one list (List LF) of words of low frequency. The words used in the construction of the lists were the same as those in the S lists and R lists of Exp. I (see Table 1). Two different pairings of the words in each list were used equally often. For purposes of verbal discrimination learning, one of the items in each pair was designated as correct. Each item in a pair was correct for half the *Ss*.

Procedure.—The members of a pair were presented one above the other in the window of a Hull-type memory drum. The rate of presentation was 1.5 sec. for the pair alone, and 1.5 sec. for the pair and the correct R together. The intertrial interval was 3 sec. Four different orders of presentation were used to minimize serial learning. In each order half of the correct words appeared in the upper position and half in the lower position. The position of the correct item in each pair varied from trial to trial. Each of the four orders was used as a starting order equally often. The *Ss* began anticipating the correct R on the first trial. Learning was to a criterion of one perfect trial.

Retention was tested by relearning either 30 sec. or 7 days after the end of practice. Relearning was for five trials or to criterion, whichever took the longer.

Subjects.—With three types of lists and two retention intervals, there were six groups of 16 *Ss* each. All *Ss* were undergraduates who had English as their native language; they were not necessarily naive to rote-learning experiments but had had no prior experience with verbal discrimination learning. The *Ss* were assigned to conditions in blocks of 6, with 1 *S* per block for each combination of lists and retention intervals. The running order within each block and the assignment to different starting orders were determined by tables of random numbers. No *Ss* were lost because of failure to learn.

Results

Original learning.—There were no significant differences between the 30-sec. and 7-day groups learning a given kind of list. The mean numbers of trials to criterion for the com-

bined groups were as follows: HF—6.4 (SD = 4.0); MF—4.5 (SD = 1.5); LF—5.3 (SD = 2.5). The effects of word frequency on acquisition parallel those obtained on the S side in paired-associate learning in Exp. I, i.e., List MF is learned faster than Lists HF and LF. The variation in criterion scores is significant (F = 3.63, df = 2/93, .02 < P < .05).

Recall and relearning.—Figure 2 shows the mean numbers of correct Rs on the five trials of relearning 30 sec. and 7 days after the end of practice. The 30-sec. test reflects the rank order of the conditions in original learning. Immediate relearning proceeds less steadily for List HF than for the other two lists. The disadvantage of List HF becomes pronounced on the 7-day test. A noteworthy fact is that the difference between List HF and the other two lists increases between Trial 1 and Trial 2 and that this separation is maintained through Trial 4. The conditions which slowed up the acquisition of List HF rapidly come into play again in relearning. The separation between Lists MF and LF is consistently small.

The differences on the recall trial are not significant. For the five trials of relearning, the over-all variation among lists is significant (F = 7.52, df = 2/90, P < .01). With performance on List HF poorest on both tests, the Time × Lists interaction falls short of significance (F = 2.12, df = 2/90). The pattern of differences and the results of the statistical tests are the same when trials to relearn to criterion are considered.

Discussion

The results of the experiments are consistent with the assumption that unit-sequence interference and unit-sequence facilitation covary as a function of word frequency, with the relative amount of effective interference increasing more rapidly on the S than on the R side. Such systematic changes in the pattern of transfer effects are indicated by the findings on acquisition in Exp. I: (*a*) rate of overt errors is a positively accelerated function of the word frequency of both Ss and Rs; (*b*) speed of learning increases steadily with the word frequency of Rs; (*c*) speed of learning first increases and then decreases as a function of the word frequency of Ss.

Speed of learning is influenced by the word frequency of both Ss and Rs whereas the amount of retention varies significantly only with S conditions. In acquisition the availability of the R units, which is a function of frequency, is a major determinant of the speed of learning (Underwood & Schulz, 1960). Once the Rs are available to the learner, the associates elicited by the Ss are of major importance in determining the speed at which the prescribed associative connections are established. Retention losses occur as incorrect associations which had been unlearned during acquisition recover in strength relative to the prescribed sequences. On the recall test, the competing Rs elicited by the S terms largely determine the amount of effective interference. The rate at which such competing associations recover with the passage of time reflects the balance of positive and negative transfer effects in acquisition. There appears to be no differential decline in the availability of the Rs when well integrated units such as words are used.

Experiment II shows that the pre-experimental probabilities of specific associates have a major influence on learning and retention when the word frequency of the Ss is low but have no differential effects when the word frequency of the Ss is high. In the acquisition of List S_h, strong associative connections either existed or were readily established by direct mediational links between any pair of words in the list, and the primary associates had no advantage over associates of low rank. In the case of List S_l, on the other hand,

the uncontrolled transfer effects of pre-experimental habits were weak and did not mask the difference between associates of high and low probability.

The results of the retention tests in Exp. II clearly bring out the differences in unit-sequence interference and facilitation falling on the pairs in the two lists. When facilitation by pre-experimental habits obtains for the pairs in both lists (S_h-A_h and S_l-A_h), retention is inversely related to the word frequency of the Ss. Under these conditions the rate of forgetting directly reflects the assumed difference in unit-sequence interference. In the case of the low-probability pairs (S_h-A_l and S_l-A_l) there is interference from language habits in both lists, as indicated by the intralist errors in List S_h and the frequent importation of primary associates into List S_l. For these pairs, retention varies directly with word frequency, now reflecting the assumed difference in unit-sequence facilitation. The question arises, of course, whether the differences in recall, and for that matter in acquisition, are simply the result of a set to give primary associates. While such a set was undoubtedly present, the critical finding is that it was effective only with Ss of low word frequency. With Ss of high word frequency, no one associate is clearly dominant, and Rs from outside the list cannot compete effectively with those within the list.

The fact that speed of verbal discrimination learning first increases and then decreases as a function of word frequency (Exp. III) provides additional support for the assumed covariation of unit-sequence facilitation and interference. With homogeneous pairs such as were used in this study, differential effects of R availability are minimized (cf. Runquist & Freeman, 1960), and the relationship parallels that obtained for Ss in paired-associate learning. The availability of strong differential Rs is essential for the achievement of a stable discrimination, but once the number of such Rs exceeds an optimal value, generalization across items counteracts the beneficial effects of R-produced differentiation. It is likely that generalization both within pairs and between pairs contributes to the total effect. While recall was too high to yield reliable differences, List HF is clearly inferior to the other two lists in relearning.

In agreement with the results obtained earlier with serial lists (Postman, 1961), the present experiments offer no support for the assumption that meaningfulness necessarily favors retention. By maximizing either the negative or positive transfer effects of pre-experimental habits, it is possible to obtain either an inverse or a direct relationship between meaningfulness and retention. For an average sample of verbal materials, the best prediction at present is that the differences in retention will be small when degree of original learning is held constant.

Summary

Three experiments investigated the transfer effects of language habits on the acquisition and retention of verbal associations. They tested the assumption that both positive and negative transfer effects (unit-sequence facilitation and unit-sequence interference) increase as a function of the frequency of usage of words. The balance of unit-sequence interference and facilitation determines the speed of acquisition and the rate at which interferences recover with the passage of time.

In Exp. I lists of paired associates representing all possible combinations of S and R terms of high, medium, and low word frequency were used. Speed of acquisition first increased and then decreased as a function of the word frequency of Ss but varied directly with the word frequency of Rs. Amount of retention did not vary significantly as a function of R conditions whereas the effect of S frequency paralleled that obtained in acquisition. While the availability of Rs and pre-experimental associative probabilities both influence speed of acquisition, amount of retention loss appears to be determined primarily by the recovery of competing associations elicited by the S terms.

In Exp. II associates of high and low pre-experimental probability, as determined by norms of free association, were learned to *Ss* of high and low word frequency. Pre-experimental associative probability has significant effects on learning and retention when the

word frequency of the Ss is low but not when it is high. Whether or not manipulation of the conditions of facilitation and interference is successful depends on the magnitude of the uncontrolled transfer effects of pre-experimental habits.

The method of verbal discrimination learning was used in Exp. III with pairs of words of either high, medium, or low word frequency. Speed of acquisition first increased and then decreased as a function of the word frequency of the paired items. Thus, the relationship paralleled that obtained for the word frequency of Ss in paired-associate learning. The differences in recall were unreliable, but the high-frequency lists were relearned more slowly than those of medium or low frequency.

References

CIEUTAT, V. J., STOCKWELL, F. E., & NOBLE, C. E. The interaction of ability and amount of practice with stimulus and response meaningfulness (m, m') in paired-associate learning. *J. exp. Psychol.*, 1958, **56**, 193–202.

HUNT, R. G. Meaningfulness and articulation of stimulus and response in paired-associate learning and stimulus recall. *J. exp. Psychol.*, 1959, **57**, 262–267.

KIMBLE, G. A., & DUFORT, R. H. Meaningfulness and isolation as factors in verbal learning. *J. exp. Psychol.*, 1955, **50**, 361–368.

L'ABATE, L. Manifest anxiety and the learning of syllables with different associative values. *Amer. J. Psychol.*, 1959, **72**, 107–110.

MANDLER, G., & CAMPBELL, E. H. Effect of variation in associative frequency of stimulus and response members on paired-associate learning. *J. exp. Psychol.*, 1957, **54**, 269–273.

MORIKAWA, Y. Functions of stimulus and response in paired-associate verbal learning. *Psychologia*, 1959, **43**, 437–446.

POSTMAN, L. Extra-experimental interference and the retention of words. *J. exp. Psychol.*, 1961, **61**, 97–110.

RUNQUIST, W. N., & FREEMAN, M. Roles of association value and syllable familiarization in verbal discrimination learning. *J. exp. Psychol.*, 1960, **59**, 396–401.

THORNDIKE, E. L., & LORGE, I. *The teacher's wordbook of 30,000 words.* New York: Teachers College, Columbia University, 1944.

UNDERWOOD, B. J., & POSTMAN, L. Extra-experimental sources of interference in forgetting. *Psychol. Rev.*, 1960, **67**, 73–95.

UNDERWOOD, B. J., & SCHULZ, R. W. *Meaningfulness and verbal learning.* Philadelphia: Lippincott, 1960.

(Received May 11, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 1, 20-23

CONDITIONED DIMINUTION OF THE UNCONDITIONED RESPONSE AS A FUNCTION OF THE NUMBER OF REINFORCEMENTS[1]

H. D. KIMMEL AND H. S. PENNYPACKER

University of Florida *Duke University*

Kimble (1961) has recently shown that the amplitude of the unconditioned eyeblink is attenuated by the presence of the CS during reinforced CS-UCS conditioning trials. Assuming that this diminution of the UCR is a manifestation of a conditioned inhibitory process under the control of the CS, Kimble and Ost (1961) varied the CS-UCS interval during conditioning to determine whether or not the degree of inhibition varied as a function of the interstimulus interval. They found that the strength of the conditioned inhibitory process was related to the interstimulus interval in the same way as is the conditioned excitatory process, i.e., the greatest amount of recovery in UCR amplitude occasioned by the omission of the CS occurred when the interval during conditioning had been 0.5 sec.

The purpose of the present study was to extend these eyelid conditioning findings to another classically conditionable human response, the GSR, and to investigate the relationship between this conditioned inhibitory process and another parameter of the conditioning process, the number of reinforcements. It was hypothesized that attenuation of the UCR during reinforcement and recovery of the UCR upon omission of the CS would both vary positively with variations in the number of reinforcements.

[1] The junior author's collaboration in this study was supported by a predoctoral research fellowship from the National Institute of Mental Health (Mf 13,327).

This study was done while the senior author was Visiting Assistant Professor at Duke University on leave of absence from the University of Florida.

METHOD

Subjects.—Thirty students in introductory psychology at Duke University were *Ss*. The *Ss* volunteered to meet a class requirement. They were assigned randomly to three groups of 10 each, to receive different numbers of reinforcements, but were otherwise treated identically.

Instructions.—The *S* read the instructions from a typewritten card. They indicated that this was an experiment on the effect of environmental stimulation upon the GSR and that *S*'s task was to remain relaxed and motionless but to pay attention to the stimuli during the experiment.

Apparatus.—The CS was a 1000-cps tone produced by General Radio Company equipment and delivered to *S* via Trimm Company earphones. The intensity of the tone, rated at the earphones, was 40 db. (re.: .0002 dynes/cm^2).

The UCS was an electric shock produced by a Psychological Instruments Company stimulator and delivered to the index and middle fingertips of *S*'s right hand. The shock intensity used during conditioning was 2 ma. at 50 v., for a hypothetical *S* with a resistance of 25,000 ohms.

The durations of the stimuli were controlled electronically. The CS had a duration of 1.0 sec. and the UCS a duration of 0.1 sec. On paired conditioning trials a delayed paradigm was used with a 0.9-sec. interval between onsets of the two stimuli.

The GSR was measured as the maximum decrease in resistance which occurred within 3 sec. after the offset of a stimulus. The response was picked up from the palm and back of *S*'s left hand by ½-in. zinc electrodes, covered with a few drops of zinc sulphate solution, in lucite cups filled with saline electrode jelly. It was amplified by a Hunter GSR apparatus and recorded on an Esterline-Angus ink-writing milliammeter with a paper

speed of 3 in/min. All responses were transformed to the square root of conductance change ($\sqrt{\Delta C}$) for statistical purposes.

Procedure.—Data were collected in a quiet dark room. An electric fan masked extraneous sounds. The *E* and the apparatus were located in an adjoining room.

The *S*'s left hand was cleaned with acetone before the GSR electrodes were attached. He was then seated in the experimental room and given the instructions to read before the pickup leads were attached to the GSR electrodes, the shock electrodes were taped to his fingertips, and the earphones were placed on his head.

The *S* was first given 2 presentations of the CS alone followed by 3 presentations of the shock alone, in intensity increments up to the intensity used during conditioning. This was followed by 4 additional presentations of the CS alone. During the reinforcement period, which occurred next, the three groups of *S*s received either 4, 8, or 16 paired presentations of the CS and UCS. After reinforcement 2 presentations of the shock alone were given and, finally, 4 presentations of the CS alone (i.e., extinction). All presentations (i.e., trials) were separated by intervals of 30 to 60 sec., varied unsystematically by *E*.

Results

Of primary significance in evaluating the hypotheses was the way in which the amplitude of the UCR varied over trials during reinforcement and the degree to which an increment in amplitude of the UCR was produced by the omission of the CS after reinforcement. Figure 1 shows the average amplitude of GSR for each group of *S*s, as a function of reinforced trials. Also indicated in Fig. 1 are the average amplitudes of GSR on the two test trials of the UCS only (shown in the figure). Each data point represents the mean of 10 *S*s.

Figure 1 indicates that the amplitude of the UCR gradually reduced during reinforcement, although the groups differed in the extent to which this reduction occurred. Statistical analysis of the degree of diminution of GSR amplitude was achieved by

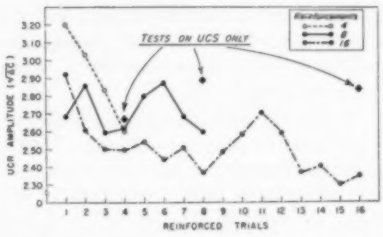

FIG. 1. Average amplitude of the unconditioned GSR during reinforcement and on test trials without the CS.

comparing the average amplitude of GSR on the first two reinforced trials (combined) with the average amplitude of GSR on the last two reinforced trials (combined). Thus, for the group which received four reinforcements the two values used were the average of the first two reinforced trials and the average of the third and fourth reinforced trials. The values used for the group which received eight reinforcements were obtained from the first two reinforced trials and the seventh and eighth reinforced trials, etc. Analysis of variance of these average measures of UCR amplitude during reinforcement showed that the overall reduction in UCR amplitude was statistically significant ($P < .05$, error variance = .2616) but that it did not interact significantly with the number of reinforcements as had been expected.

Figure 1 also shows the degree to which the omission of the CS produced an increment in UCR amplitude. The exact quantitative nature of the relationship between the number of reinforcements and the degree of increment produced by the omission of the CS is shown in Fig. 2, which presents the average difference (i.e., increment) in GSR amplitude between the last two reinforced trials (combined) and the two test trials of UCS only (combined). A log-

arithmic scale has been used for the abscissa in Fig. 2 to emphasize the linearity of the relationship when expressed in this way.

It is clear in Fig. 2 that the degree to which the omission of the CS produced an increment in UCR amplitude increased as a linear function of the logarithm of the number of reinforcements, i.e., as a negatively accelerated growth curve. The analysis of variance of the averaged UCRs on the last two reinforced trials and the two UCS-only trials indicated that the overall increment in UCR amplitude produced by the omission of the CS was statistically significant ($P < .025$) and that the interaction between this increment and the number of reinforcements approached significance ($P < .07$). The linear component of the trend over log N was statistically significant ($P < .025$). Error variance for these tests = .1331.

It was of interest also to examine the relationship between the number of reinforcements and GSR amplitude during extinction. This evaluation was contaminated in part by the fact that all of the *S*s had received two presentations of the UCS alone between the end of reinforcement and the beginning of extinction, which probably tended to obscure any differences in degree of conditioning which might have occurred. For this reason only the last three extinction trials were used to evaluate the degree of conditioning which occurred. When the average amplitude of GSR on these three trials was compared to the average amplitude of the GSR on the last three presentations of the CS alone prior to conditioning it was found that the average amplitude of GSR increased greatly from before to after the reinforcement period and that this increase was largest in the group which received eight reinforcements. The overall difference from before to after reinforcement was highly significant, supporting the conclusion that conditioning occurred, but the interaction between the number of reinforcements and degree of conditioning failed to reach significance.

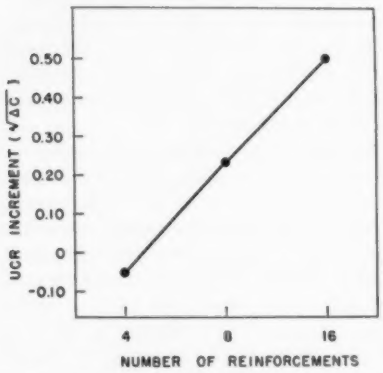

FIG. 2. The relationship between the number of reinforcements and the amount of increment produced by the omission of the CS.

DISCUSSION

The results of this experiment support the notion that the CS becomes capable during conditioning of attenuating the amplitude of the UCR. Not only did the amplitude of the UCR diminish during reinforcement, but, when the CS was omitted and the UCS presented alone, the UCR recovered to approximately its original amplitude.

It is of little significance to point out that *S* was surprised to receive the UCS without the CS and that this surprise, acting as a startle stimulus, may have been responsible for the increment in the GSR on shock-alone trials. The fact that the presence of the CS serves as a warning to *S* that the shock is coming is, of course, one of the things that are "learned" during the conditioning process. Even when viewed in this way, the reduction-of-surprise function of

the CS is essentially an inhibitory function. That this inhibitory aspect of the classical conditioning process is highly adaptive to the organism being conditioned is apparent, since a good deal of unnecessary emotional overresponding is brought under control or avoided by the association of the warning signal (i.e., the CS) with the noxious event which is imminent.

The relationship between the present findings and those of an earlier study of disinhibition in GSR conditioning (Kimmel & Fowler, 1961) is of some significance. In the earlier study it was shown that the addition of an extra, or disinhibiting, stimulus to the CS after reinforcement had the effect of producing an increment in the amplitude of the conditioned GSR and that this effect was positively related to the number of reinforcements given prior to the addition of the extra stimulus. The interpretation of the results of the earlier study was based upon the assumption that the CS acquired inhibitory capabilities during reinforcement and that the extra stimulus inhibited this inhibition, producing an increment in response amplitude, i.e., disinhibition. The present findings support this interpretation, since the omission of the CS had the same behavioral consequence as the inhibition of the inhibition controlled by the CS, i.e., in both cases the inhibition was removed or reduced. The observed positive relationship between these effects and the number of reinforcements supports the contention that they result from an associative process as, of course, do the findings of Kimble and Ost (1961) that the conventional interstimulus interval function is involved.

Summary

This study tested the hypotheses that the amplitude of the unconditioned GSR is gradually attenuated by the CS during conditioning and that this reduction in UCR amplitude vanishes when the CS is omitted and the UCS presented alone. Three groups of 10 *Ss* received either 4, 8, or 16 reinforced presentations of a tone CS and electric shock UCS in a classical delayed paradigm. Following reinforcement the CS was omitted and 2 UCS alone trials were administered. After the tests on the shock alone, 4 extinction trials of the CS alone were given.

It was found that the amplitude of the unconditioned GSR diminished during reinforcement and that the UCR recovered to approximately its original amplitude when the CS was omitted. The amount of increment produced by omitting the CS was a linear function of the logarithm of the number of reinforcements. These results were interpreted as supporting the notion that a conditioned inhibitory process develops during reinforcement, under the control of the CS, which attenuates the amplitude of the UCR in the presence of the CS. The relationship between these findings and similar findings in a study of the relationship between amount of disinhibition and number of reinforcements was thought to add further support to the interpretation offered.

References

Kimble, G. A. *Hilgard and Marquis' conditioning and learning.* New York: Appleton-Century-Crofts, 1961.

Kimble, G. A., & Ost, J. W. P. A conditioned inhibitory process in eyelid conditioning. *J. exp. Psychol.*, 1961, **61**, 150–156.

Kimmel, H. D., & Fowler, R. L. The relationship between disinhibition resulting from compound stimulus presentation and the number of reinforcements. *Off. Naval Res. tech. Rep.*, 1961, No. 3. (Contract NONR 580—09)

(Received May 15, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 1, 24–28

THE EFFECTS OF DRIVE AND DISCRIMINATION TRAINING ON STIMULUS GENERALIZATION [1]

DAVID R. THOMAS [2]

Duke University

In 1959 Hanson reported in this journal a study on the influence of discrimination training on stimulus generalization. Hanson trained pigeons to peck at a key illuminated by a monochromatic light of 550 mμ, under a VI reinforcement schedule. After successive discrimination training, *S*s were subjected to a generalization test in extinction, following a procedure introduced by Guttman and Kalish (1956).

Hanson found that time to the discrimination criterion varied as a negatively accelerated decreasing function (S+, S−) difference. Post-discrimination generalization gradients (PDGs) were higher and steeper than a control gradient and in addition, showed a shift in the peak of responding from the S+ in the direction away from the S−. The extent of this shift varied also as a negatively accelerated decreasing function of (S+, S−) difference.

The present study is an extension of Hanson's work. The *S*s were tested on three different discrimination problems at three different levels of food deprivation. The literature is inconsistent with regard to the influence of drive level on discrimination learning, and no study appears to have been reported using the operant free-responding technique. In this study, *S*s were tested for generalization before discrimination training was initiated, and in addition, short generalization tests were interspersed during the course of discrimination training so as to trace the development of the changes in the PDG.

METHOD

Subjects.—The *S*s were 54 experimentally naive white carneau pigeons obtained from the Palmetto Pigeon Plant in Sumpter, South Carolina.

Apparatus.—A bank of four identical Skinner type key pecking apparatuses was used. Each box had the following internal dimensions: width, 15½ in.; depth, 14½ in.; height, 14½ in. Walls and ceiling were painted flat black; floors were of unpainted Masonite. The *S*'s key of translucent plastic was exposed through a ¾-in. circular aperture placed 6½ in. above the floor on one side of the box. Directly below the key in the floor of each box was a 1-in. circular aperture through which *S*s had access to the food magazine. The magazine, which was operated by a motor-driven cam, allowed *S* approximately 4 sec. of access to food during each cycle. Between cycles, the food was lowered beyond *S*'s reach. A Plexiglas light fixture with a 15-w. bulb, placed directly above and in front of the floor opening illuminated the opening and the magazine for the duration of each cycle.

Aside from the stimulus light and the magazine light during magazine cycles, the boxes were in darkness throughout the experiment. One box, Number 1, was set aside for generalization testing. The source of illumination for the key in this box was a Bausch and Lomb diffraction grating monochromator, Model 33-86-40, equipped with a 108-w., 6-v., ac ribbon filament tungsten lamp.

Boxes 2, 3, and 4 were used for training

[1] This study was taken from a dissertation submitted to the Psychology Department of Duke University in partial fulfillment of the requirements for the PhD degree in psychology. The writer is greatly indebted to Norman Guttman under whose direction the investigation was conducted. Thanks are also due to Doris Homa Thomas for running some of the *S*s. At the time the research was conducted the author was a Predoctoral Research Fellow of the United States Public Health Service.

[2] Now at Kent State University.

purposes only. The keys of these were illuminated by 100-w., 120-v., ac projection lamp. Positive and negative stimuli were provided by Bausch and Lomb interference filters. Each box was equipped with a "positive filter," (transmission peak 550 mμ) and a "negative filter," (590 mμ for Box 2, 570 mμ for Box 3, and 558 mμ for Box 4). The brightness of the filter colors was matched to the monochromator colors by the addition to the filters of an appropriate number of pieces of exposed film (Polaroid No. D 77843). The matching was accomplished with the aid of a photomultiplier tube and its associated amplifier.

The three training boxes were set up to change stimuli automatically according to a prearranged schedule punched on a tape run on a Gerbrands programer. Throughout the entire experiment, masking noises were supplied to all boxes by a Grason-Stadler noise generator, Model 901.

Procedure.—Upon arrival at the laboratory, all *Ss* were weighed, individually caged, and allowed free access to food and water. Throughout the entire experiment free access to water was always available in the home cages. After 4 to 10 days of free feeding, a stable weight level was achieved by all *Ss* and food deprivation was begun. The *Ss* were randomly assigned to three body weight levels, 80% of ad lib. weight, 70%, and 60%, and to the three different discrimination problems, thus creating nine groups of *Ss*, one for each combination of weight level and discrimination problem. Deprivation ceased for each *S* when the appropriate weight was reached. At this point training was begun.

All *Ss* were adapted, magazine trained, and conditioned to key peck according to a set schedule covering 5 days. During this time, the box was in complete darkness with the exception of the key light and the magazine light during reinforcement cycles.

All *Ss* were given VI reinforcement for 5 days following conditioning. Thirty 60-sec. stimulus-on periods, alternated with 10-sec. stimulus-off periods, were given each day. The mean interval between reinforcements, not counting stimulus-off periods, was approximately 60 sec, with a range from 4 sec. to 4 min. During stimulus-off periods, a shutter operated, removing the light from the key and leaving the experimental box dark. Reinforcements were never given during a stimulus-off period. One of the five VI sessions for each *S* was administered in Box 1, the test box, so as to accustom *Ss* to the test box and to the monochromator-produced stimulus.

On the next day, after a 5-min. warm-up period of VI in the test box, each *S* was subjected to a generalization test in extinction. Eleven different test stimuli were used—510, 520, 530, 540, 545, 550, 555, 560, 570, 580, and 590 mμ. The 11 test stimuli were randomized within a series and six different random series were presented to each *S*. This resulted in a schedule of 66 stimulus presentations. Each stimulus presentation was for 30 sec. and was followed by a 10-sec. blackout period, during which the number of responses in the preceding period was recorded and the stimulus was changed.

On the day following the (preliminary) generalization test, discrimination training was begun. All three discrimination groups were trained with the same positive stimulus (S+) 550 mμ, and differed only with respect to the negative stimulus (S−). Each discrimination group consisted of 18 *Ss*, 6 at each of the three body weight levels. These groups may be designated by the negative stimulus used: 590 mμ (trained in Box 2), 570 mμ (Box 3), and 558 mμ (Box 4). During discrimination training, responding to the positive stimulus was reinforced according to the same VI schedule previously used. Responding to the negative stimulus was never reinforced. The positive and negative stimuli were presented successively in a prearranged random order. Fifteen 1-min. intervals of S+ and 15 of S− were presented each day. All stimulus changes were made during the 10-sec. blackout periods. The 30 stimulus presentations comprised three blocks of 10, and within each block there were 5 positive and 5 negative stimuli. Discrimination training was continued until a criterion of no responding in five successive periods of S− combined with continued responding to S+ was achieved.

During the course of discrimination training, a three-series generalization test of 33 stimulus presentations was administered at the completion of every even numbered daily session of discrimination training. If the discrimination criterion was met on a day on which an interpolated generalization test was scheduled, the test was omitted. On the day following the achieving of the criterion for discrimination, a final generalization test was administered to each *S*. The final test was carried out in the same manner as the preliminary and interpolated tests, with the exception that 12 test series (132 stimulus presentations) were used.

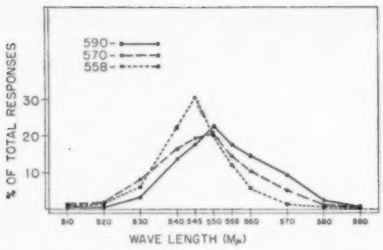

FIG. 1. Postdiscrimination generalization gradients after three different discrimination problems, pooled over three drive levels.

RESULTS AND DISCUSSION

Data bearing on three different problems will be presented in this report. The first problem concerns the roles of hunger drive level and (S+, S−) difference in determining the rate of formation of a discrimination. The present experiment is fully in agreement with Hanson (1959) in the finding that the rate of formation of a discrimination varies inversely with the physical difference between the stimuli to be discriminated. Analysis of variance indicates that minutes to criterion varies as a function of the (S+, S−) difference ($F = 23.50$, $df = 2/45$, $P < .01$). A beneficial influence of drive level is suggested, but is not demonstrated at a statistically acceptable level of confidence ($F = 2.96$, $df = 2/45$, $.05 < P < .10$). The results seem sufficiently promising, however, to warrant a more direct attack of the problem, without the interpolation (as in the present case) of other experimental treatments (preliminary and interpolated generalization tests). The interaction between the (S+, S−) difference and drive level is not significant ($F < 1$, $df = 4/45$).

In the present experiment deprivation levels down to 60% of ad lib. weight were employed. Thus, the limits of the hunger motive were assessed. We may therefore conclude that even at its extremes, the hunger drive exercises far less control over the rate of discrimination learning than does the physical difference between the stimuli to be discriminated.

The second problem to be considered here concerns the roles of drive level and the (S+, S−) difference in determining the location of the PDGs. In Fig. 1 are presented the PDGs of the three problem groups pooled over all drive levels. The PDGs show the characteristic displacement of the peak from the S+ an amount inversely related to the (S+, S−) difference. A convenient measure of the location of a gradient is derived by treating it as a grouped frequency distribution and computing the mean value. The difference between the means of the 18 individual mean scores in the three problem groups is significant ($F = 10.74$, $df = 2/45$, $P < .01$). The effect of the (S+, S−) difference on the location of the PDG is mirrored at all three levels of drive. However, the effect of drive level on the means of the final generalization gradients fails to achieve a statistically acceptable level of confidence ($F = 2.56$, $df = 2/45$, $.05 < P < .10$). Neither is the interaction between the (S+, S−) difference and drive level significant ($F = 1.07$, $df = 4/45$). Thus, Hanson's findings with regard to the effect of the (S+, S−) difference on the location of the central tendency of the PDG are replicated at three different levels of drive.

In agreement with Hanson, we have found that both the time to criterion and the amount of displacement of the PDG are inversely related to the (S+, S−) difference. This suggests the possibility that amount of training is the vehicle through which the (S+, S−) difference has its effect on the PDG. The analysis of this ques-

tion was the third major problem to which the present study addressed itself. The administration of a short generalization test to all *S*s after 2 days of training (with the exception of those *S*s which had already met the learning criterion) makes possible an evaluation of the effects of the (S+, S−) difference, with amount of training held constant. On the other hand, the analysis of the series of generalization tests given to all *S*s during the course of discrimination training reveals the role of amount of training with the (S+, S−) difference controlled.

Although only 1 *S* from the 590-mμ group failed to achieve the criterion of discrimination learning within 2 days, 14 *S*s from the 570-mμ group and 16 *S*s from the 558-mμ group could be compared in this manner. For each *S* a "displacement score" was obtained by subtracting the mean of the second generalization test from the mean of the first. The means of these displacement scores for the two groups were virtually identical, 5.6 mμ for the 570-mμ group vs. 5.7 mμ for the 558-mμ group. The mean displacement is the same in spite of the fact that in one case the (S+, S−) difference is more than twice that of the other!

TABLE 1

DISPLACEMENT OF THE MEAN OF THE GENERALIZATION GRADIENT WITH TWO SESSIONS OF DISCRIMINATION TRAINING AT TWO DIFFERENT (S+, S−) VALUES

570-mμ Group		558-mμ Group	
S No.	Displacement of Mean (mμ)	*S* No.	Displacement of Mean (mμ)
2	12.37	15	7.91
5	2.39	54	−3.11
14	7.77	66	4.55
56	14.72	51	7.90
70	−0.62	12	6.49
46	6.59	18	8.42
47	7.04	44	9.52
59	4.77	48	2.40
72	−0.97	60	4.25
43	2.22	67	7.91
40	3.21	32	−2.25
37	0.12	38	2.75
31	7.82	25	9.64
62	10.27	22	14.44
		29	7.07
		68	3.37
Mean	5.6	Mean	5.7

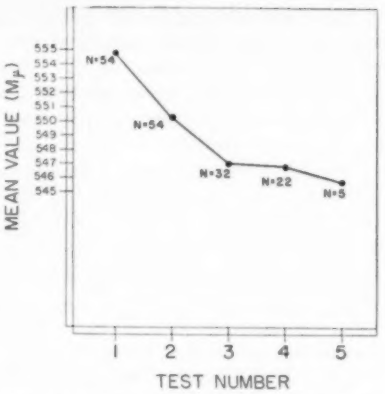

FIG. 2. Mean of the gradient as a function of ordinal position of test.

In Fig. 2 are presented the means of the preliminary generalization gradients obtained from all *S*s, the means of the second obtained gradient, the third, etc. The data from all groups have been pooled since it appears that neither the (S+, S−) difference nor drive level have any marked effect on the location of the PDG if the amount of training is held constant. Under each of the values is recorded the number of *S*s whose scores are represented in that mean value. It should be noted that the points plotted were not obtained from different groups of *S*s. On the contrary, the 5 *S*s which received five generalization tests were included among the 22 *S*s which received (at least) four generalization tests, which,

in turn were among the 32 *S*s which received (at least) three generalization tests, etc. The value of *N* continually decreases indicating that all *S*s had at least two generalization tests whereas progressively fewer *S*s had three tests, four tests, etc. The figure shows a negatively accelerated decreasing function. Even as few as two sessions of discrimination training produces a marked shift in the location of the gradient.

The present procedure has made possible the separation of the effects of the (S+, S−) difference and amount of training on the location of the PDG. When this is done it appears that the effects of the (S+, S−) difference are mediated through the amount of training. There is no obvious explanation of the failure of the (S+, S−) difference to have a *direct* effect on the location of the PDG. It is clear, however, that this finding raises serious doubt about the adequacy of a Spence-type theory of inhibition (Spence, 1936, 1937) to account for the process of successive discrimination learning in the operant free-responding situation. It is to be hoped that future studies will suggest an alternative theoretical framework which will have greater predictive and explanatory value within this context.

Summary

The effects of drive and discrimination training on stimulus generalization were studied in the pigeon. Three groups of 18 *S*s each, maintained at 60%, 70%, and 80% of ad lib. weight, respectively, were trained to respond to a key illuminated by a light of 550 mμ. After 5 days of VI training, *S*s were tested for generalization to stimuli ranging from 500 mμ to 600 mμ. Then the three groups were subdivided into three discrimination problem groups, 550 mμ positive for each, but with 590 mμ, 570 mμ, and 558 mμ negative, respectively. During discrimination training, responding to S+ was VI reinforced, responding to S− was never reinforced. Periodically during the course of discrimination training short generalization tests were given to all *S*s. After the criterion of discrimination was met, all *S*s were subjected to a final generalization test.

The major conclusions were: (*a*) minutes to criterion varies inversely with the (S+, S−) difference; a beneficial effect of drive is suggested but not conclusively demonstrated ($.05 < P < .10$). (*b*) Discrimination training produces a general steepening of the PDG, a lowering of the gradient in the region of S−, and a shift of the central tendency from the region of S+ in the direction away from S−, the amount of shift varying inversely with the (S+, S−) difference. This finding is replicated for all three levels of drive. (*c*) Generalization gradients obtained during the course of discrimination training reveal that the mean of the gradient shifts in a negatively accelerated manner as a function of the amount of discrimination training. The amount of training appears to be the vehicle through which the effect of the (S+, S−) difference on the location of the PDG is mediated. Displacement varies with amount of discrimination training, independent of any direct effect of the (S+, S−) difference.

References

Guttman, N., & Kalish, H. I. Discriminability and stimulus generalization. *J. exp. Psychol.*, 1956, **51**, 79–88.

Hanson, H. Effects of discrimination training on stimulus generalization. *J. exp. Psychol.*, 1959, **58**, 321–334.

Spence, K. W. The nature of discrimination learning in animals. *Psychol. Rev.*, 1936, **43**, 427–449.

Spence, K. W. The differential response in animals to stimuli varying within a single dimension. *Psychol. Rev.*, 1937, **44**, 430–444.

(Received May 19, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 1, 29–34

THE LEARNING OF RESPONSES TO MULTIPLE WEIGHTED CUES[1]

STANLEY A. SUMMERS[2]

University of California, Los Angeles

In many situations a response is dependent simultaneously on several different cues, i.e., on several potentially useful stimulus attributes. The purpose of this experiment was to analyze the learning of responses to simultaneously presented cues of different validities, in order to determine how much the responses came to depend on each cue. The implications are relevant especially to situations in which individuals react to several influences in reaching a judgment, decision, or interpretation.

Questions about the effects of complex stimuli have been answered by several related types of experimental analysis, such as investigations of stimulus-compound patterning, of probability learning, and of the utilization of multiple cues. In studies of cue utilization, correlations may be taken between each of several cues and the response variable. This correlational approach, which is particularly associated with the work of Brunswik (1956), was used in the present experiment. Brunswik developed the concepts of ecological validity and functional validity. Ecological validity is the correlation between the cue and an environmental variable predicted by it. Functional validity is the correlation between the cue and a response, such as a judgment, directed toward the environmental variable. Stimulus attributes, then, can be dealt with as joint cues to some criterion variable. A correlational approach to cue utilization provides a method of abstracting from a series of responses the simultaneous effects of several cues.

Brunswik assumed that the functional validity of a cue will correspond, in most cases, to its ecological validity; the extent to which each cue is used should come, through learning, to conform to the extent to which the cue predicts. With correlational analysis Brunswik investigated, for the most part, established patterns of cue utilization, rather than the learning process by which individuals acquire such patterns. He attempted to ascertain the use of cues in settings representative of the natural environment. To isolate the effect of cue validity, however, it seems desirable to establish ecological validities for originally meaningless cues in a laboratory setting and to observe the responses learned to these cues.

An attempt to study the learning of functional validities in a controlled situation was made by R. Goodnow (Bruner, Goodnow, & Austin, 1956). The validity of a cue for predicting a correct category was defined as the relative frequency of the cue's association with that category. The results did not indicate any close correspondence between cue use and objective validity. However, there are limitations in the treatment of this question in terms of frequencies rather than correlations. One limitation is that probability matching, which was considered the criterion of appropriate functional validity, would not have maximized success in the situation and, therefore, did not have the adaptive character of a match of functional to ecological validity in a correla-

[1] This paper is based upon a PhD dissertation presented to the Graduate School of the University of California, Los Angeles. The author wishes to express his gratitude to Wendell E. Jeffrey for guidance and assistance throughout the investigation. He also wishes to thank Joseph A. Gengerelli and Irving Maltzman for their suggestions.

[2] Now at San Fernando Valley State College.

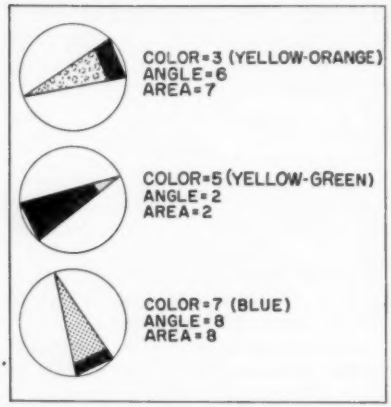

FIG. 1. Three stimulus forms.

tional framework. Smedslund (1955), dealing less directly with the relation between ecological and functional validity, approached the topic of cue learning through a correlational framework. He established joint visual cues to a predicted variable. A correlation was imposed between each of the cues and the correct response, giving each cue a different validity. The main conclusion was that *S*s do learn "to utilize many probabilistic cues simultaneously . . ." (Smedslund, 1955, p. 26). In Smedslund's design, however, a particular cue was always associated with the same validity, so that the effect of validity could not be studied independent of the effects of saliency or other cue characteristics.

In the present study, correlations were imposed between each of three simultaneously presented visual cues and a variable, correct line length, whose magnitude varied with the magnitude of all three cues. Each of the experimentally imposed correlations will be called a *cue weighting*. Correlations were taken between correct line length and *S*'s response throughout a learning session, and between each of the individual cues and *S*'s response. These empirically derived correlations will be called *response weightings*.

It was expected that the order of utilization of the three cues would come to conform to the order of cue validity. Specifically, it was predicted that the response weightings would be different and would be ranked in the order of the cue weightings, and that the magnitude of the response weightings would approach that of the cue weightings.

METHOD

Subjects.—The *S*s were 30 members of a ninth grade class, 20 boys and 10 girls. This was a special class composed of students of high academic performance.

Materials.—The stimulus materials were 384 geometrical forms photographed on 35-mm. Kodachrome slides and projected on a screen from an automatic projector. Each slide bore a number giving its position in the standard order of presentation. The projected form was approximately 2 ft. in diameter. Each form consisted of an isosceles triangle on a circular white background. A portion of the triangle was black; the remainder was of some hue. Three characteristics of these stimuli varied from slide to slide. These were color, angle, and area. That is, the hue of the triangle, its orientation with respect to the vertical, and the proportion of the triangle that was colored rather than black. Each of the three characteristics varied over eight values; these were designated 1 through 8 (see Fig. 1).

A hue of red was assigned Value 1. The other hues and their values were red-orange (2), yellow-orange (3), yellow (4), yellow-green (5), green (6), blue (7), and violet (8). The two equal sides of the triangle were the long sides; the ratio of long to short was 5:2. If the point opposite the short side is regarded as an indicator, the orientation of the triangle can be described in terms of degrees of clockwise rotation from the 0° (12 o'clock) position. Values were assigned to these positions as follows: 22.5° (1), 67.5° (2), 112.5° (3), 157.5° (4), 202.5° (5), 247.5° (6), 292.5° (7), 337.5° (8). The triangles varied along the dimension of area from almost completely black (1), in equal increments of altitude, to almost completely colored (8). The black strip always paralleled the short side of the triangle. The values (1 through 8) of the

three dimensions varied independently of one another. A standard stimulus series was presented to all *S*s.

The response materials consisted of booklets of paper on which were printed horizontal lines approximately 7 in. long. Each line was an elongated arrow, pointing toward the right. The first 2 pages of a booklet contained 64 lines, numbered 1 to 64, corresponding to numbers on the first 64 slides. The next 16 pages contained 256 bracketed pairs of lines, each pair given a number corresponding to the number in the slide sequence. The lower line of each pair was completely covered by a cardboard strip, which was lightly pasted at both ends. Hidden by this strip was a small red mark which designated a correct length for that line. The last 2 pages of the booklet were identical to the first 2.

Procedure and design.—All *S*s were tested in a single session of approximately 2 hr. Two *E*s were present throughout the experiment. The instructions were a fairly straight-forward description of the task and its significance. The following is an abstract of the instructions.

> The experiment will investigate how we learn to make certain kinds of judgments. You will see a series of slides projected on a screen—like this. Each of these forms contains the information you need to tell you how to mark off a particular length along an arrow on answer sheets like this. You will just make a small dash, measuring from the left, for how long you think the line should be. Sometimes it will go here, sometimes there. Each slide contains all the information you need to mark off the correct length. Your task in the experiment is to learn how to use this information. At first you will get no help in estimating how long the line should be. Later you will be able to find out how long the line should be after making your own guess. The answer sheets for most of the experiment will have two arrows attached like this. The correct answers are covered by paper strips. You can make a mark right on the line over the strip and then pull down or tear back the strip to see the correct answer. Then you can compare this answer with the mark you made and with the picture. Be careful not to look at the correct mark until after you make your own. Each design on the slide tells you how long the line should be by the features that change from slide to slide. Which way the figure points, which color is on the slide, and the amount of color are all clues to how long the line should be, and just how, you will have to learn during the experiment. Using all these features at the same time you may learn to get the line length just right. This is difficult but the closer you can come the better. The first two pages are a preliminary measurement before you get help in estimating how long the line should be. The last two pages, also, do not have correct answers marked. You will see each slide only a short time so you will have to judge pretty much by your first quick impression.

A series of 384 slides was presented; each slide was projected for 10 sec., with a 2-sec. interval between slides. During each exposure *S* viewed the slide, made a mark on the upper line of the corresponding line-pair, tore back the cardboard strip, and compared the correct line length marked on the lower line with his mark and with the stimulus pattern. For the first 64 and the last 64 trials, where no correct length was given, *S* simply viewed the stimulus and marked the line. This provided a measure of operant level and an extinction series. A 10-min. rest period was given in the middle of the presentation.

The stimulus series of 384 figures was composed of six blocks of 64 each. The last block (E) was identical to the first (O). Within each block of 64, each of the eight values of each cue appeared equally often in combination with each value of the other two cues. Each block was based on one of five different latin squares, selected from a complete set of 8×8 latin squares (Fisher & Yates, 1948, p. 63). The effect of this arrangement was to provide a zero correlation between any two cues within each block and to make almost all stimuli different, aside from the extinction series. Within each block the order of the various combinations of the three cues was randomized, and the distribution of values for each cue was rectangular.

The relation of the cue values to the corresponding series of correct line lengths was defined in the following way. Correct line length was made a function of the three cue values, as expressed by the equation $CLL = 2C_1 + 1.5C_2 + C_3$, where CLL is correct line length, and the Cs are values of the three cues. The coefficients of the Cs determine the cue weightings. Since the minimum cue value was 1, and the maximum 8, correct line length varied from 4.5, when all cue values were 1, to 36, when all cues had the value 8. The relation between each cue and correct line length can also be expressed in terms of a correlation. Since the cues were uncorrelated, the correlations of the three cues

TABLE 1

EXPERIMENTAL CONDITIONS: WEIGHTINGS ASSIGNED EACH CUE IN DETERMINING CORRECT LINE LENGTH

Condition	Cue Type and Weighting		
	.74	.56	.37
1	Color	Angle	Area
2	Color	Area	Angle
3	Angle	Color	Area
4	Angle	Area	Color
5	Area	Color	Angle
6	Area	Angle	Color

with correct line length were in the ratio 2:1.5:1. Since the square of a correlation corresponds to the proportion of total variance predicted, it is possible to calculate the value of the correlation coefficient for a cue by the equation:

$$(2r)^2 + (1.5r)^2 + r^2 = 1.00$$
$$r = .371$$

Thus the three correlations, or cue weightings, were .371, .557, and .743. The procedures provided *S* with all the information necessary to make a correct response if he used each cue correctly.

In order to ascertain the effect of various cue weightings independent of other properties of the cue, six experimental conditions were provided. The design was completely counterbalanced with respect to cue type (color, angle, area) and cue weighting and their combinations (see Table 1). These six conditions were six different sets of correct line lengths for the single series of slides. The procedure for the various conditions differed only in the construction of the response booklet, that is, in the correct line lengths. All six experimental groups participated in the experiment together, being presented with the stimulus series simultaneously. The *Ss* were randomly assigned to experimental conditions. Data were obtained from 33 *Ss*; data from 1 randomly chosen *S* in each of three conditions were discarded to provide 5 *Ss* in each condition.

RESULTS

For each block of 64 presentations, four Pearson correlation coefficients were computed for each *S*. These were the correlations between his response and each cue, as well as between his response and correct line length. The correlations were averaged over *Ss* by converting each correlation to Fisher's *z*, averaging the *z*'s, and converting the average to a Pearson *r*. Curves showing average response weightings for each cue weighting are presented in Fig. 2. The vertical position of the curves corresponds to the order of the cue weightings.

An analysis of variance was performed on *z* transformations of the response weightings for the learning blocks, the four blocks in which correct line length was presented to *S*. The main question to be answered by the analysis is whether *Ss* were learning to respond differentially to the different cue weightings; that is, whether there was a significant difference in vertical placement of the curves. The analysis (Lindquist, 1953, Type VI) is summarized in Table 2. The *F* for cue weighting is significant beyond the .001 level. This result confirms the prediction that the response weightings would differ with the cue weightings.

In addition, there is an over-all trend toward increasing cue utilization over these four blocks. The block effect is significant beyond the

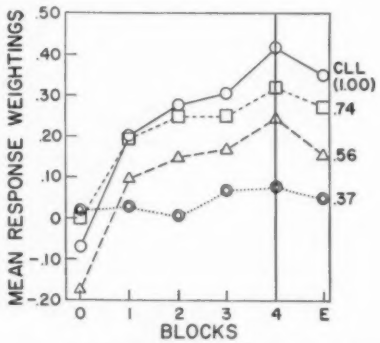

FIG. 2. Mean response weightings (correlations of cue with response) for different cue weightings. (Correct line length was presented in Blocks 1 through 4.)

TABLE 2

ANALYSIS OF VARIANCE OF RESPONSE WEIGHTINGS USING A z TRANSFORMATION

Source	*df*	*MS*	*F*
Between *Ss*			
Conditions (C)	5	.1382	<1
Error (b)	24	.2529	
Within *Ss*			
Cue weighting (W)	2	1.3726	8.96**
Blocks (B)	3	.1858	5.40*
W × B	6	.0233	1.03
W × C	10	.2071	1.35
B × C	15	.0154	<1
W × B × C	30	.0276	1.22
Error (w)			
$Error_1$	72	.0344	
$Error_2$	48	.1531	
$Error_3$	144	.0227	

* $P < .005$.
** $P < .001$.

.005 level. In order to ascertain whether there is a trend toward increasing utilization of each separate cue, the block effect was tested for each cue weighting separately. The block effect was significant for the .74 weighting ($P < .05$, $F = 2.79$, $df = 3/72$) and for the .56 weighting ($P < .025$, $F = 3.34$, $df = 3/72$), but not for the .37 weighting ($F < 1$). The form of the learning curves does not differ for different cue weightings; the W × B interaction is not significant.

The question of the final magnitudes of the response weightings could not be answered since there was no indication that learning had reached a limit. It is possible, however, to take the observed correlation of response with correct line length on a given block and to compute what the three response weightings would be if they contributed to this total correlation in the expected ratio 2:1.5:1. The response weightings computed on this basis are presented in Table 3, along with the corresponding observed weightings. Table 3 indicates that cue utilization, throughout the learning trials, is roughly proportional to cue validity, although the highest-weighted cue is being used slightly more, and the lowest slightly less, than would be predicted on the basis of proportionality.

DISCUSSION

The results indicate that *Ss* responded simultaneously and differentially to the multiple cues. The differential response is shown by the highly significant cue weighting effect. Simultaneous response to the multiple cues, rather than use by each *S* of only a single cue, is indicated by the fact that this weighting effect transcends the $S \times$ Weighting interaction contained in the error term ($Error_2$). Also, once *Ss* are exposed to correct line length, the correlation of response with correct line length is higher than with the most heavily weighted cue (see Fig. 2), indicating successful use of more than the most valid cue by itself. Individual *Ss* must have responded to more than one cue during any given block of trials. The curves suggest that there is some response to the lowest-

TABLE 3

OBSERVED MEAN RESPONSE WEIGHTINGS (O) COMPARED WITH RESPONSE WEIGHTINGS COMPUTED ON THE BASIS OF PROPORTIONALITY (P)

Cue Weighting	Response Weighting			
	Block			
	1	2	3	4
CLL	.203	.271	.313	.418
.74				
P	.15	.20	.23	.31
O	.19	.24	.25	.32
.56				
P	.11	.15	.17	.23
O	.10	.15	.17	.24
.37				
P	.08	.10	.12	.16
O	.04	.01	.06	.07

weighted cue, although in the statistical test this was not rising significantly. In general the *Ss* made an appropriate use of the cues available.

The importance to adaptive behavior of appropriate response to multiple cues has been emphasized for various situations: in terms of the learning of functional relations (Miller, 1959), for prediction of judgmental behavior (Johnson, 1955), with application to clinical appraisal (Hammond, 1955; Hoffman, 1960). Important differences of emphasis are implied in the choice of any one methodological framework for the study of cue utilization. In particular one may contrast the orientation implied by a correlational framework with that implied by an event frequency framework, such as is used in studies of probability matching. Both frameworks may be used to study adaptation to environmental uncertainties. The event frequency situation provides for the success of a given response to be probabilistic. Usually the best adaptation possible is consistent choice of the most frequent alternative, with probability matching implying less than optimal adjustment. In a correlational framework, the probabilistic character of the individual cues does not necessarily adhere to the cues in combination, which may be unequivocally related to the criterion as they were in the present study. In terms of behavior, the matching of response weighting to cue weighting would imply fully optimal adjustment to the situation; there may be consistently accurate responses on the basis of individually limited cues. A correlational framework, then, emphasizes cue utilization as a process which tends to remove uncertainty. Much of our judgmental and discriminative behavior is characterized by a sufficient degree of sureness and adaptiveness to be best studied within such a framework.

Summary

This experiment investigated the relation between the objective validity of certain cues and the extent to which these cues were used. The purpose was to analyze the learning of responses to multiple cues of different validities, in order to determine how much the responses came to depend on each cue. The independent variable was the correlation imposed between a criterion and each of three simultaneously presented visual cues. The dependent variable was the correlation between the cues and *Ss*' responses.

The *Ss* were 30 members of a ninth grade class. The stimulus materials were geometric forms. Three characteristics of these stimuli varied, acting as cues for the prediction of a line length. These characteristics were based on the color, angle, and area of parts of the forms, and were counterbalanced to establish independent cue validities. The cue validities of .74, .56, and .37 together permitted a perfect prediction of line length. During four blocks of 64 trials each, *Ss* responded and then were presented with the correct line length.

Successful cue utilization increased during the learning session. The *Ss* responded to different cues simultaneously, and the extent to which cues were used differed with validity. Cue utilization was roughly proportional to cue validity throughout the learning trials.

References

Bruner, J. S., Goodnow, J. J., & Austin, G. A. *A study of thinking.* New York: Wiley, 1956.

Brunswik, E. *Perception and the representative design of psychological experiments.* Berkeley: Univer. California Press, 1956.

Fisher, R. A., & Yates, F. *Statistical tables for biological, agricultural, and medical research.* London: Oliver & Boyd, 1948.

Hammond, K. R. Probabilistic functioning and the clinical method. *Psychol. Rev.*, 1955, **62**, 255–262.

Hoffman, P. J. The paramorphic representation of clinical judgment. *Psychol. Bull.*, 1960, **57**, 116–131.

Johnson, D. M. *The psychology of thought and judgment.* New York: Harper, 1955.

Lindquist, E. F. *Design and analysis of experiments.* New York: Houghton Mifflin, 1953.

Miller, N. E. Liberalization of basic S-R concepts. In S. Koch (Ed.), *Psychology: A study of a science.* Vol. 2. *General systematic formulations, learning, and special processes.* New York: McGraw-Hill, 1959.

Smedslund, J. *Multiple-probability learning.* Oslo: Akademisk Forlag, 1955.

(Received May 23, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 1, 35–39

STIMULUS-RESPONSE CONTIGUITY IN CLASSICAL AVERSIVE CONDITIONING [1]

R. A. CHAMPION

University of Sydney

Stimulus-response contiguity is widely accepted as a necessary, if not sufficient, condition of learning. In particular, it has been suggested that even though reinforcement may be required for complete learning, a response which is being learned cannot be reinforced until it has begun to occur and that S-R contiguity may operate as a crucial factor in *initiating* any learning process (Mason, 1959). In classical conditioning, degree of S-R contiguity is represented in the temporal interval between the CS and the UCR, and the following experiment was designed to test the hypothesis that, early in conditioning, performance is proportional to degree of CS-UCR contiguity. In order to make this test a short-latency response (eyeblink) and a long-latency response (GSR) were conditioned simultaneously in each *S* and the time relations between CS and UCR were varied so as to permit greater contiguity of eyeblink and CS in one group of *S*s and greater contiguity of GSR and CS in a second group. This type of arrangement is represented in Fig. 1; with an air puff as the UCS, the CS (tone, T) may be presented near the unconditioned eyeblink as in the case of T1, or near the unconditioned GSR as with T2. In terms of the contiguity hypothesis it was predicted that the eyeblink would condition better than the GSR with T1 whereas the reverse would hold with T2, i.e., that there would be interaction between the CS-UCS interval and the latency of the UCR as independent variables affecting conditioning. By the same token it was expected that the early conditioning of the eyeblink would be better with T1 than with T2 but that T2 would produce the superior performance in the case of the GSR.

The experiment provided an opportunity to confirm an earlier finding that backward conditioning produced some true learning of the long-latency GSR (Champion & Jones, 1961). It is apparent from Fig. 1 that UCS-CS trials must be used in the conditioning of the GSR if optimum CS-UCR contiguity is to be achieved and this arrangement was therefore employed in the second of the two groups described above. After these two groups had been trained, however, the question arose as to the real meaning of "backward conditioning." The accepted definition seems to be put simply in terms of the presentation of the UCS before the CS, but this does not guarantee that the UCR will precede the CS. In fact, in the second

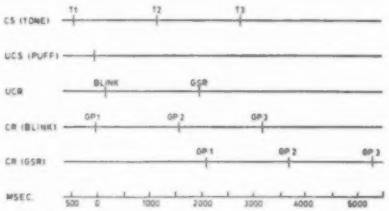

FIG. 1. Schematic representation of time relations between stimuli and responses in the three groups of the experiment. (T1, T2, and T3 represent the location of the CS in Groups 1, 2, and 3, respectively.)

[1] Part of the apparatus used in this study was provided by Commonwealth of Australia Research Grant No. 2002.

arrangement described above, favoring contiguity of GSR and tone (T2 in Fig. 1), the CS still precedes the UCR. A third group was therefore trained with an even longer backward UCS-CS interval so as to ensure that the unconditioned GSR preceded the CS (T3 in Fig. 1). An attempt was made to set this interval so that degree of contiguity was nevertheless approximately the same as in the second group, i.e., so that CS-UCR contiguity matched UCR-CS contiguity. The use of this third group allowed a test of the possibility that R-S and S-R contiguity are equally effective.

Method

Subjects.—The *S*s were 49 volunteers from courses in psychology at the University of Sydney.

Apparatus.—The general experimental situation and the apparatus for measuring the GSR have been described elsewhere (Champion & Jones, 1961). The eyeblink was measured with the system developed at the State University of Iowa laboratory, i.e., a microtorque potentiometer linked mechanically to *S*'s right eyelid and electrically to an ink-writing recorder through a dc amplifier. The CS was a 2000-cps, 90-db. tone of 50-msec. duration delivered to *S* through headphones. The UCS was a 3.0-psi air puff applied to *S*'s right eye through a .062-in. diameter tube and set at 50 msec. duration by means of an ac solenoid valve. All time intervals except the intertrial interval were controlled with electronic timers.

Procedure.—In the course of general instructions *S*s were asked to keep their eyes open except for normal blinking. The series of trials for each *S* began with one UCS-alone trial followed by one CS-alone trial. The remaining presentations of stimuli consisted of training trials (CS-UCS or UCS-CS) interspersed with test trials (CS alone) in the following order: 3 training, one test, 7 training, one test, 10 training, and one test trial. Thus test trials to assess the progress of conditioning were administered after 0, 3, 10, and 20 training trials. The training series was limited to 20 trials because the hypothesis under test dealt only with the early stages of conditioning. The intertrial interval averaged 30 sec. and varied between 25, 30, and 35 sec. in prearranged random order. No ready signal was given at any stage.

As represented in Fig. 1, Groups 1 and 2 were conditioned with a CS-UCS interval of 400 msec. and a UCS-CS interval of 1200 msec., respectively. Group 3 was then trained with a UCS-CS interval of 2800 msec. A conditioned eyeblink was defined as occurring on test trials with a pen deflection of 1 mm. or more within the interval 200–500 msec. following the onset of the CS. No attempt was made to exclude "voluntary responders." A GSR on test trials was taken to be any response occurring in the interval 1–4 sec. after the CS. There was some adaptation of the GSR to the air puff and the results of 4 *S*s who failed to give a response on 6 or more of the 20 training trials were discarded. This left 15 *S*s in each of the three groups. There were 7, 8, and 6 women in Groups 1, 2, and 3, respectively; a survey of the results showed that the men gave insignificantly more CRs of each type compared with the women.

Results

In view of the predicted interaction between the length of the CS-UCS interval and the latency of the UCR it was necessary to make a direct comparison between the performance of the eyeblink and the GSR. For this purpose a technique was used which had already proved effective (Mason, 1959) and which allowed a measure of the frequency of the conditioned GSR. For each *S* the GSR scores (change in conductance in micromhos) obtained on test trials during training were expressed as a ratio of the response on the first test trial, before training; this transformation overcame individual differences in general sensitivity of response. In the few cases where there was no response on the first test trial the least measurable response of 0.1 micromhos was assumed. The transformed scores, ranging from 0 to 35.0, were pooled and found to have a median of 2.6. Each value at or above 2.6 in the transformed scores was then taken as a conditioned GSR.

The eyeblink and the GSR were conditioned simultaneously in each *S* and it was possible to test for the predicted interaction by comparing across groups the proportion of *S*s giving more conditioned eyeblinks than conditioned GSRs on the test trials during training. These data are set out in Table 1; the result for Groups 1 and 2 was in accord with the prediction and a Fisher exact probability test (Siegel, 1956), applied with cases of equal frequency of response omitted, showed the interaction to be significant ($P < .01$ for a two-tailed test). Comparison of these results with the data from Group 3 showed the interaction to be significant for Groups 1 and 3 ($P < .05$) but not for Groups 2 and 3.

TABLE 1

NUMBER OF *S*S GIVING MORE, EQUAL, AND FEWER CONDITIONED EYEBLINKS COMPARED WITH CONDITIONED GSRS ON TEST TRIALS DURING TRAINING

Outcome (CRs)	Group 1 (N = 15)	Group 2 (N = 15)	Group 3 (N = 15)
Eyeblinks > GSRs	9	0	3
Eyeblinks = GSRs	3	4	2
Eyeblinks < GSRs	3	11	10

The other aspect of the predicted interaction involved within-response comparisons over the various CS-UCS intervals. Performance curves for the eyeblink CR are shown in Fig. 2. The performance of Group 1 was consistently superior to that of Group 2, as predicted, but the application of Fisher tests showed that a statistically significant difference did not emerge until 20 training trials ($P < .01$); at 10 trials the difference approached significance ($.10 > P > .05$). The same relationship held between Groups 1 and 3 but there was no significant difference at any stage between Groups 2 and 3. An overall test at 20 trials proved significant ($\chi^2 = 12.09$, $P < .01$, $df = 2$). These results conform to the general finding that a 400-msec. interval is almost optimum for the eyeblink and that backward conditioning of this response is ineffective.

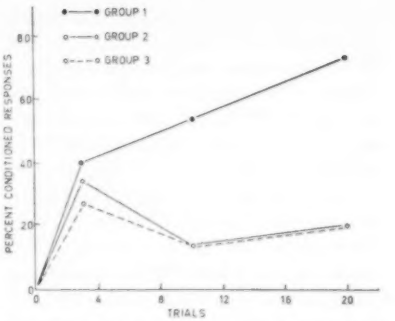

FIG. 2. Performance curves for eyeblink conditioning.

The more important within-response comparison concerned the GSR, and performance curves for the conditioning of this response are shown in Fig. 3. As predicted, Group 2 was consistently superior to Group 1, but according to Fisher tests the difference was only statistically significant at 3 trials ($P < .05$). Groups 1 and 3 did not differ significantly at any stage. An

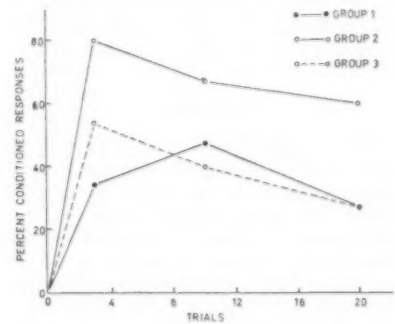

FIG. 3. Performance curves for GSR conditioning.

overall test at 3 trials gave a significant result ($\chi^2 = 6.66$, $P < .05$, $df = 2$). More sensitive measures than frequency of response were available for the GSR and the transformed amplitude scores were studied for further information. Nonparametric statistical tests were used because of the marked positive skew in the distribution of these scores. The results given above were confirmed; in addition, the application of a U test at 20 trials showed that the difference between Groups 1 and 2 approached significance ($U = 68$, $.10 > P > .05$). At this stage there was a significant difference in amplitude of response between Groups 2 and 3, with the former superior ($U = 51$, $P < .05$), but not between Groups 1 and 3. An overall test applied to the three groups at 20 trials proved significant ($H = 6.99$, $P < .05$, $df = 2$).

As a check on the possible introduction of some artifact in the transformation of the GSR scores, Mood's test of trend (McNemar, 1955) was applied to the raw change-in-conductance scores for each group. The observed trends in these data were virtually identical with those in the frequency data, depicted in Fig. 3; they were significant for Groups 2 ($\chi^2 = 11.8$, $P < .01$, $df = 3$) and 3 ($\chi^2 = 11.4$, $P < .01$), but not for Group 1 ($\chi^2 = 2.20$, $P > .50$).

Discussion

The predictions about the effects of S-R contiguity in classical conditioning were confirmed by the data. It may be asked why exact contiguity of CS and UCR was not chosen as the condition most likely to yield optimum performance early in conditioning; with the assumption of latencies of 200 msec. and 2000 msec. for the unconditioned eyeblink and GSR, respectively, this would have been achieved with the use of corresponding backward UCS-CS intervals. A pilot study conducted along these lines before the present experiment produced results exactly in keeping with the contiguity hypothesis, but the level of conditioning in both groups was so low that the interaction was not statistically significant. Poor conditioning with a short backward interval is an established result for a short-latency response (e.g., Wolfle, 1932) and the results of Groups 2 and 3 in the present experiment suggest that there is a decrease in the effectiveness of GSR conditioning in early stages with a backward interval somewhere between 1200 msec. and 2800 msec., possibly of the order of 2000 msec. An obvious inference to be drawn from these considerations is that the required contiguity lies not between CS and UCR as overt events but rather between corresponding physiological processes, lagging behind the CS in stimulus reception and preceding the UCR in response evocation.

It is tempting to suppose that the CS must precede the UCR for contiguity to be effective. This would explain why the conditioning of the eyeblink did not occur in Groups 2 and 3, and why Group 3 was inferior to Group 2 and comparable with Group 1 on some counts in the conditioning of the GSR. The supposition is contradicted, however, by the comparable performance of Groups 2 and 3 with respect to the relative frequency of the two types of CR (Table 1) and by the presence of a significant trend in the GSR conditioning of Group 3. Both these latter results are probably due to the early rise in the performance of Group 3, at three training trials. Contiguity appears to have had some effect when the UCR preceded the CS (Group 3), but the performance was inferior to and not as sustained as that obtained with a comparable degree of CS-UCR contiguity (Group 2). Such an outcome could be due to two factors which are not mutually exclusive. First, if allowance is made for the physiological processes

delineated above, then the contiguity between stimulus reception and response initiation would be greater in Group 2 than in Group 3 (Fig. 1); under the conditions of Group 3, of course, some physiological mechanism must be found which permits contiguity to act when response initiation precedes stimulus reception. Second, if delay of reinforcement is represented in the interval between CR and UCS, then it is greater in Group 3 than in Group 2 and performance in the former group should not have been so sustained as in the latter group for that reason; this consideration brings with it the problem of "backward" reinforcement when the UCS precedes the CR. More generally, it may seem improper to invoke the effects of reinforcement if attention is limited to the initial stages of learning, where contiguity has been taken as the prime factor operating. Reinforcement may take effect, however, as soon as the learned response appears, no matter how weakly, and the precise separation of this effect from that of contiguity awaits the use of some more refined technique than mere reference to early and late stages of learning.

Attention should be drawn to two other features of the data. There is some conflict with the results of the previous study (Champion & Jones, 1961), where the backward conditioning of the GSR (with a UCS-CS interval of 500 msec.) was inferior from the outset to forward conditioning (with a CS-UCS interval of 500 msec.) whereas the performance of Group 2 was superior to that of Group 1 in the present study. The discrepancy might be accounted for in terms of differential degrees of contiguity and reinforcement, but the differences in procedure in the two experiments are too great to permit this explanation to be pursued with any confidence. A second point of interest was the appearance in the present data of a statistically significant contiguity effect early in training with the GSR but relatively later with the eyeblink. This result would emerge if greater contiguity was achieved by chance in Group 2 with the GSR than in Group 1 with the eyeblink. It should be noted that the only set of conditions favoring performance on the grounds of both contiguity and reinforcement was that prevailing in Group 1 for the eyeblink and that only under these conditions did the performance curve rise throughout its course.

Summary

A test was made of the hypothesis that CS-UCR contiguity is an important factor in the initial stages of classical aversive conditioning. The short-latency eyeblink and long-latency GSR were conditioned simultaneously in each *S*, with tone as CS and air puff as UCS. A forward CS-UCS interval of 400 msec. and a backward UCS-CS interval of 1200 msec. were used in separate groups to favor contiguity of CS with eyeblink and GSR, respectively. The acquisition of the eyeblink was superior with the forward interval and inferior with the backward interval when compared with that of the GSR. Within-response comparisons showed that the eyeblink conditioned better with the forward interval whereas the GSR conditioned better with the backward interval. To test the effect of UCR-CS contiguity in the case of the GSR a third group was trained with a backward UCS-CS interval of 2800 msec.; this condition produced an initial rise in performance which was not sustained. The results of the experiment were interpreted as supporting the hypothesis.

References

Champion, R. A., & Jones, J. E. Forward, backward, and pseudoconditioning of the GSR. *J. exp. Psychol.*, 1961, **62**, 58–62.

McNemar, Q. *Psychological statistics.* (2nd ed.) New York: Wiley, 1955.

Mason, J. E. The joint action of contiguity and reinforcement in classical conditioning. Unpublished doctoral dissertation, University of Sydney, 1959.

Siegel, S. *Nonparametric statistics.* New York: McGraw-Hill, 1956.

Wolfle, H. M. Conditioning as a function of the interval between the conditioned and the original stimulus. *J. gen. Psychol.*, 1932, **7**, 80–103.

(Received May 27, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 1, 40–45

ASSOCIATIONS, SETS, AND THE SOLUTION OF WORD PROBLEMS [1]

MIRIAM A. SAFREN [2]

Johns Hopkins University

Investigators have used word problems (anagrams and skeleton words) to explore the "category set," which can be defined as a readiness to respond to words belonging to a common category or class, and can be measured by the speed or frequency[3] with which certain responses occur. For example, Starch (1911) and others demonstrated the existence of the category set when *Ss* solved problems made from lists of words belonging to a common class more rapidly than problems made from random word lists. Later, Rees and Israel (1935) and others measured the strength of the category set by the frequency with which *Ss* solved multiple solution problems as category words after practice on unique solution problems which could be solved only as words of the selected category. Past *Es* reported that telling *Ss* the category name did not always have a consistent effect; what they are surer of is that the set arises from verbal context during problem solving. Unfortunately, the past *Es* selected the category words and category name on an a priori basis, and none of them systematically investigated the verbal context triggering the set. The present study aims to show that associative strength between words from which problems are made is a pertinent variable to manipulate in the operation of the category set in anagram solving.

This approach stems from work in free recall showing that associative strength between words presented to *Ss* significantly determines *Ss'* verbal productions. Additional support for this approach is given by Mayzner and Tresselt (1958) who suggest that processes in anagram solving may be described by laws similar to those in verbal recall.

Especially relevant to the problem investigated here is a study by Deese (1959) who investigated the effect on recall of presenting *Ss* with lists of words all of which are associated, with the average strength of association among the group expressed in an index he called interitem associative strength. He found that in "organized" lists of words high on this index, *Ss* recalled more and had fewer verbal intrusions. Postulating that free recall is in part free associating, Deese explained that *Ss* recalled more with the organized lists because there was a greater probability of *Ss'* associations actually being listed items.

The present study extends Deese's model of free recall to anagram solving and hypothesizes that the category set in anagram solving operates by associations between words selected by *Es* as belonging to a given category. On the basis of this hypothesis,

[1] This study is based upon a dissertation submitted to the Johns Hopkins University in partial fulfillment of the requirements for the PhD degree, and was supported in part by Public Health Service Fellowship Number MF 13,450 from the National Institute of Mental Health. The author expresses her thanks to James E. Deese for his guidance throughout the study.

[2] Personnel Research Psychologist, Bureau of Old-Age and Survivors Insurance, Social Security Administration, Department of Health, Education, and Welfare.

[3] The probability of occurrence of response can be considered the definition of set. The probability of occurrence of a response is then inferred from the time it takes the response to occur or the frequency with which it occurs.

TABLE 1

THE SIX ORGANIZED LISTS USED IN THE EXPERIMENT

List 1			List 2			List 3		
Word	Anagram	Labels (Cond. O-L)	Word	Anagram	Labels (Cond. O-L)	Word	Anagram	Labels (Cond. O-L)
COMMAND	CANDOMM	Military	WHISTLE	WHELIST	Railroad	CHAIR	RAICH	Comfort
ORDER	REDOR	Armed Forces	TRAIN	TARNI	Station	SOFT	FOST	Relaxation
ARMY	RAMY	Discipline	NOISE	NESOI	Sounds	SOFA	OFAS	Furniture
OBEY	OYEB		SOUND	UNDOS	Noises	CUSHION	CUNSHIO	
SOLDIER	SODLERI		SHRILL	SHLIRL		PILLOW	WOLPIL	
NAVY	VANY		LOUD	OLDU		COUCH	COCHU	
List 4			**List 5**			**List 6**		
MILK	LIMK	BEVERAGES	DOCTOR	DOROCT	Hospital	SQUARE	SAQURE	Geometry
CREAM	RECAM	Breakfast	NURSE	URNSE	Physical Health	CIRCLE	CIRLEC	Geometric
SUGAR	USGRA	Food	HEALTH	EHALHT	Medical Care	ROUND	DUNRO	Figures
COFFEE	EFECOF	Taste	SICK	KISC		CUBE	BECU	Shapes
SWEET	SEWTE		MEDICINE	MEDCIENI		BLOCK	CLOBK	
DRINK	DINRK		CURE	ECRU		BALL	LALB	

the following predictions were made: (*a*) Anagrams made from organized lists of associatively related words will be solved more quickly than anagrams made from random word lists, because in the former, associations called up by solved problems should aid *S*s in the solution of subsequent ones. (*b*) There should be a rapid decrease in solution time with trials when *S*s solve anagrams made from associatively related words as associations called up by solved problems *begin* to aid *S*s in the solution of subsequent ones.

METHOD

The study consisted of two parts: the selection of stimulus materials and the experiment itself.

The experiment.—The experiment consisted of three major conditions. In Cond. O and Cond. O-L groups of *S*s were presented with anagrams made from "organized" lists of words (words that elicit each other in free association). In addition, *S*s assigned to Cond. O-L were told labels describing the list of words from which the anagrams were made. These labels were selected on the basis of normative data gathered prior to the experiment. The labels were associates to the words of the list and in turn elicited them. In Cond. R, *S*s were given anagrams made from "random" lists of words (words with essentially zero probability of eliciting each other in free association). The experimental design was such that the same sample of words in the same anagram form was used for all experimental conditions. Thus, as many factors as possible were constant for all conditions and whatever effects might occur could be attributed to the influence of associative context.

Stimulus materials.—The stimulus materials consisted of 36 anagrams. The 36 words from which the anagrams were made came from 6 organized lists of 6 words each. These lists were constructed after administering several word association tests consisting of 200 items including filler items to groups of from 50 to 100 Johns Hopkins University undergraduates. The instructions given to *S*s in this part of the study were standard word association test instructions.

The following governed the selection of the six words of each organized list: (*a*) the six words should have maximum interitem associative strength, i.e., the words should have a maximum probability of eliciting each other in free association, and (*b*) they should be words from which only unique[4] solution anagrams could be made. The six lists of six words each are shown in Table 1.

Labels describing the organized lists (Cond. O-L) were obtained as follows: A

[4] Two of these words do not form unique solution anagrams. These are SOFA which could be structured as OAFS and CIRCLE which could also be CLERIC. In the anagram form in which they were presented, none of these solutions occurred.

TABLE 2

OVER-ALL MEDIAN SOLUTION TIMES (IN SEC.) FOR SIX ORGANIZED LISTS OF DIFFERING INTERITEM ASSOCIATIVE STRENGTHS WITH AND WITHOUT LABELS

List	Interitem Associative Strength		No. Words Labels Elicit		Median Time (No Labels)	Median Time (with Labels)
	Mean	Variance	Mean	Variance		
1	8.0	245.07	1.11	1.32	13.65	6.60
2	7.4	146.86	1.38	.96	7.65	3.85
3	6.1	79.40	1.22	.90	10.85	2.90
4	5.2	95.46	1.86	1.06	3.25	1.25
5	8.3	137.72	2.03	1.69	16.45	2.10
6	9.1	183.40	1.64	1.18	3.25	1.25

Note.—Median time scores for each list are based upon data of 6 *Ss* who solved six anagrams each.

group of 60 *Ss*, Johns Hopkins University undergraduates, were given a 6-page booklet containing the six organized lists, one per page. The *Ss* were told to look at each list of words and to think of a label or labels that described them or the way in which they were similar, and to note this in the space provided. The various labels and the frequency with which they were given were tabulated. The labels selected for each list were those which included 50% or more of the responses. The labels are shown in Table 1.

As the labels of each list were associates to the words of that list, further normative data were obtained to find out if the labels would elicit the words of the list. A sample of 40 *Ss*, Johns Hopkins University undergraduates, were given a 6-page booklet containing the six sets of labels, one per page. The *Ss* were told to look at the labels on each page and then to write down the words which the labels called to mind, but to spend no longer than 1 min. associating to each set of labels. Table 2 shows the average number of words of each organized list which the labels of that list elicited. The mean number of words called up by these labels in 1 min. was 10 (SD = 6.32).

TABLE 3

ANAGRAM SOLUTION TIMES (IN SEC.) FOR THE THREE CONDITIONS

Condition	N	Median	Q	Range
R	36	12.2	6.2	2–67.8
O	36	7.4	5.6	.1–49.2
O-L	36	2.8	1.9	.6–22.4

Note.—Time scores on which medians, Qs, and ranges are calculated are the median solution times for each of the 36 *Ss* tested under each condition.

For Cond. R in which *Ss* were presented with anagrams made from random lists of words, 36 random lists[5] of six words each were generated as follows: One word at a time was taken from each of the 6 organized lists to form a random list of six words. By a prearranged order, every word of each organized list occurred in 6 of the 36 random lists but never in the context of any other word of the same organized list.

In constructing the anagrams, an attempt was made to control for three factors that affect difficulty of solution (Mayzner & Tresselt, 1958, 1959; Sargent, 1940): (*a*) the degree of difference of the letter arrangement from the original order in the word, (*b*) the transitional probabilities between the letters in the disarranged form, and (*c*) frequency of occurrence of the word in printed matter (Thorndike & Lorge, 1944).

Subjects.—A total of 108 *Ss* from the introductory psychology class at the Johns Hopkins University were randomly assigned, 36 per condition. For Cond. O and O-L there were 6 lists of six anagrams each. Different groups of 6 *Ss* received each of the 6 lists, and within each group the order of occurrence of the six anagrams was counterbalanced over *Ss*. For Cond. R, there were 36 lists of six anagrams each. Each of the 36 *Ss* assigned to this condition received a different list.

Procedure.—All *Ss* were presented with six anagrams typed in capitals on 3 × 5 in. cards and told the following: (*a*) they would have six anagrams to solve; (*b*) an anagram is a word with the letters disarranged; (*c*) they were to look at the anagram, try to determine the original word without the use

[5] The 36 random lists are in the doctoral dissertation on file at the library of the Johns Hopkins University.

of paper and pencil, and then to give the solution orally; (*d*) they would have 4 min. to solve each problem and then would be given the solution; (*e*) they were to work as rapidly as possible as they were being timed on each problem. In the case of Cond. O-L, *Ss* were told that the anagrams were made from a list of words related by association, and given the labels that described the list. Solution times were recorded with a stopwatch, and all *Ss* were tested individually.

Results

Because of the nature of the distribution of *Ss*' anagram solution times, medians rather than means were chosen to represent *S*'s typical performance in the statistical analysis. Table 3 summarizes the data on solution time for the three conditions showing that medians, *Q*s, and ranges decrease in going from Cond. R to O to O-L. The overall statistical significance of the difference between the three experimental conditions, using the Kruskall-Wallis one-way analysis of variance, is $P < .005$. A comparison of Cond. O with Cond. R by the Mann-Whitney U test yielded $P < .02$. A similar comparison between Cond. O and O-L yielded $P < .003$.

With a significant difference between Cond. R and Cond. O, a detailed analysis was made of the organized lists of differing interitem associative strengths. A Kruskall-Wallis one-way analysis of variance by ranks showed a significant difference between organized lists ($P < .05$). However, the correlation (Kendall's tau) between interitem associative strength index and median solution time for the organized lists was $-.28$ and nonsignificant. Table 2 shows this information.

Since the comparison between Cond. O and Cond. O-L was very significant, Mann-Whitney U tests were done comparing median solution times for the 6 *Ss* of each organized list with the comparable 6 *Ss* of the same organized list with labels. For Lists 3, 5, and 6 there were significant differences between the two conditions ($P = .021$, .013, and .002, respectively). For Lists 1, 2, and 4 the differences were not significant ($P = .155$, .242, and .155, respectively). Therefore, it may be concluded that anagram solution is facilitated when *Ss* are given anagrams made from "organized" associatively related word lists. Anagram solution is facilitated still further when in addition, *Ss* are told labels describing the organized lists when these labels are (*a*) associates to the words and (*b*) in turn call up the words of the lists.

The overall Trial effect (*Ss* under all conditions had six problems to solve), as tested by the Friedman two-way analysis of variance by ranks, was significant ($P < .02$). However, there appears to be an interaction between trials and conditions (see Fig. 1), for the Trial effect is nonsignificant for Cond. R ($.50 < P < .70$) and nonsignificant for Cond. O-L ($.10 < P < .20$). The Trial effect is significant for Cond. O ($P < .05$). The significant Trial effect for Cond O verifies the predicted decrease in solution time with trials when *Ss* solve anagrams made from organized lists as associations called up by the context begin to aid *Ss* in the solution of subsequent problems.

With a significant Trial effect for Cond. O, analysis was made of the Trial effect for each organized list. Only the Trial effect for List 3 was significant ($P < .02$). However, the sample size for each organized list was only 6.

Although the six organized lists vary little in the mean probability with which the words elicit each other in free association, there is a signifi-

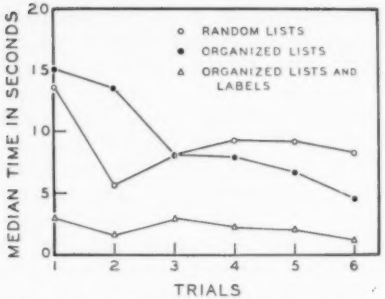

FIG. 1. Median anagram solution time per trial for each of the three experimental conditions. (Each point is based upon the data of 36 *S*s.)

cant difference in the variance about the mean (see Table 2). Bartlett's test for the difference between the K variance estimates was significant at $P < .05$. Furthermore, there seems to be a relationship between the interitem associative strength index and the decrease in solution time with trials for each organized list.[6] Ranking each list by the magnitude of its variance and then ranking each according to the amount of change in solution time with trials results in a correlation (Kendall's tau) of $-.60$, $P = .068$. This suggests that the greater the variance about the interitem associative strength index the less the change in solution time with trials.

A similar correlation (Kendall's tau) between the interitem associative strength of each list and the variance about this value is .60 ($P = .068$). Thus, it appears that organized lists with higher interitem associative strengths had a higher variance about this index which would tend to make the drop in solution time with trials less.

Possibly, not controlling for the variance about the interitem associative strength index was one reason why it was not possible to show a position relationship between the magnitude of the index and the median solution time for each organized list. It is also possible that other factors inhibiting this relationship were (*a*) the narrow range of interitem associative strengths used, and (*b*) the inability to make anagrams of all organized lists of equal difficulty.

[6] A table showing the trial effect for each organized list is in the doctoral dissertation.

DISCUSSION

The results support the hypothesis that in the solving of word problems such as anagrams, the category set operates by associations between words selected by *E* as belonging to a common category. The results also suggest processes in the solving of word problems such as anagrams may be similar to those in verbal recall; for in both verbal recall and anagram solving, associative strength between words significantly determines *S*s' verbal productions.

A suggested model to explain the results is as follows: When *S*s solve a list of anagrams, at least in part, they are calling up words to match the letters presented and sampling from a momentary response pool of available words. After *S*s have determined some of the words from which the problems were made, these words implicitly evoke others which are associatively related. These associations become part of the momentary available pool of words from which *S*s are sampling. Thus, implicitly evoked associations from the solution of preceding problems can facilitate the solution of subsequent problems by converging upon words on the list from which the anagrams were made.

When *S*s are given anagrams made from associatively related words and also given labels for these words, these labels evoke at the *outset* a response pool of available words some of which converge upon solutions to the anagrams. Since both associations elicited by solved anagrams and those called up by labels increase the availability of anagram solution, anagrams are solved most quickly under Cond. O-L. However, since from the outset the labels make

some of the anagram solutions readily available, the Trial effect is not significant for Cond. O-L.

Although couched in different language, the present model of the category set agrees basically with that of Maltzman and Morrisett (1952). The latter explain the set as follows: An anagram belonging to a common class, e.g., "nature," serves as a stimulus for the arousal of other words belonging to the selected class and in addition, a whole hierarchy of responses some of which belong to the selected class, some of which do not. However, since only the selected category responses are consistently reinforced in the experiment, these become dominant through mediated generalization.

Maltzman and Morrisett do not explicitly describe the mediated generalization mechanism. However, in the present model, the *Ss*' pool of associatively related words implicitly evoked by solved anagrams could be regarded as the mediated generalization mechanism. This pool of words would include the category name in so far as it is associated to the words comprising a common class and elicits these words.

The present description of anagram solving in terms of verbal recall and free association may be considered an extension of a model by Deese (1959), who interprets free recall as involving in part free association. Thus, the present study represents an extension over which associative relationships between words plays an important role.

The present research also implies that in future studies on the category set, stimulus materials should be calibrated by existing associative relationships for more precise predictions on the rate of problem solving. For it is the contention of this study that associative strength between words may be considered a measure of *existing response sets* which when activated, facilitate solutions of word problems such as anagrams.

Summary

This experiment tested the hypothesis that the category set in anagram solution operates by associative strength between words selected by *E* as belonging to some common class. In Cond. O and O-L, groups of *Ss* were given anagrams made from "organized" lists of words that elicited each other in free association. In Cond. O-L, *Ss* were also told labels describing the lists. The labels were associates to the words on the list from which the anagrams were made and in turn elicited these words. In Cond. R, *Ss* were given anagrams made from "random" word lists (words with essentially zero probability of eliciting each other in free association).

Time to solution was evidence for the operation of the set. It was predicted that (*a*) anagrams made from organized lists would be solved more quickly than anagrams made from random lists; for in the former, associations called up by the context would aid in problem solution; (*b*) there would be a significant Trial effect (a decrease in solution time with trials) for the organized condition as associations called up by context began to aid *Ss* in the solution of subsequent problems. The results verified these predictions and supported the present interpretation of the category set. A description of anagram solving as involving verbal recall and free association was advanced to explain the results.

References

Desse, J., Influence of inter-item associative strength upon immediate recall. *Psychol. Rep.*, 1959, **5**, 305–312.

Maltzman, I., & Morrisett, L., Jr. Different strengths of set in the solution of anagrams. *J. exp. Psychol.*, 1952, **44**, 242–246.

Mayzner, M. S., & Tresselt, M. E. Anagram solution times: A function of letter order and word frequency. *J. exp. Psychol.*, 1958, **56**, 376–379.

Mayzner, M. S., & Tresselt, M. E. Anagram solution times: A function of transition probabilities. *J. Psychol.*, 1959, **47**, 117–125.

Rees, H. J., & Israel, H. E. An investigation of the establishment of mental sets. *Psychol. Monogr.*, 1935, **46**(6, Whole No. 210).

Sargent, S. S. Thinking processes at various levels of difficulty. *Arch. Psychol., N. Y.*, 1940, **35**(Whole No. 249).

Starch, D. *Experiments in educational psychology.* New York: Macmillan, 1911.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Teachers College, Columbia University, 1944.

(Received May 27, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 1, 46–51

SEQUENTIAL INTERFERENCES DEMONSTRATED BY SERIAL RECONSTRUCTIONS[1]

E. B. COLEMAN

Human Resources Research Office

The findings of experiments on retroactive and proactive inhibition have been generalized to suggest that forgetting that occurs outside the laboratory is caused by interference similar to that found in the traditional laboratory experiment. Recently Underwood and Postman (1960) noted that "As plausible as this argument may seem, it is apparent that it would be highly desirable to specify possible sources of extraexperimental interference, and to specify these in such a way that experimental tests of their influence on forgetting are possible" (p. 74). They suggested two such interferences: "letter-sequence interferences" and "unit-sequence interferences."

Letter-sequence interferences result when a list presented for learning conflicts with English spelling habits. Suppose *S* is presented with a consonant syllable JQB. Underwood and Postman assume that previously learned spelling habits will make this sequence hard to learn because Q never follows J; nor does B ever follow Q in English spelling. These previously learned spelling habits must be "extinguished" or" unlearned" or "inhibited" before JQB can be learned. Following Briggs (1954), they also assume that as time passes, the older sequential habits will spontaneously recover and interfere with JQB during a test for retention (proactive inhibition). The older spelling sequences may be used during the delay period also (common sequences being used more frequently than uncommon ones). This will further strengthen them and further interfere with JQB (retroactive inhibition). These assumptions suggest that as time passes, *S* may replace JQB with a more common sequence—perhaps JOB.

The "unit" in unit-sequence interferences is usually the word, although it could be any sequence of letters presented as an independent unit. Word-sequence interferences result when a sequence presented for learning conflicts with syntactic habits. Suppose a list presented to *S* contains the sequence STAND DID LINGERIE THEY IN. Assumptions analogous to those given in the preceding paragraph predict that as time passes, *S* may replace such a sequence with a more familiar one—perhaps THEY DID STAND IN LINGERIE.

To collect evidence for these two sources of interference, Underwood and Postman used serial anticipation learning and measured retention after 1 week for four kinds of lists: common and uncommon words, and common and uncommon trigrams. As time passed, errors due to remembering an item in an incorrect order increased faster for word lists than for trigram lists. In a similar experiment, Postman (1961) also found more such

[1] Based on a dissertation submitted to the Faculty of Philosophy of the Johns Hopkins University in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The guidance of James Deese is most gratefully acknowledged. This research was supported in part by a National Science Foundation Fellowship (30125) and in part by a Public Health Service Research Fellowship (MF-11, 727).

intralist errors in recalling a list of common nouns than in recalling uncommon nouns. Both findings may be interpreted as evidence for word-sequence interferences. Underwood and Postman (1960) also found that as time passed, letter-sequence errors increased faster in the uncommon trigram lists than in the common trigram lists. This may be interpreted as evidence for letter-sequence interferences.

On the other hand, when the criterion is number of items recalled correctly, neither of the interferences have been revealed unless the analysis was restricted to the middle serial positions—to items that had not been overlearned. Thus evidence for the effect of these two sources of extra-experimental interference is not as strong as one would wish.

The most direct and convincing evidence for sequential interferences would be to demonstrate that uncommon sequences are replaced by more common ones during recall. This investigation will attempt to demonstrate this by a variation of serial reproductions which might be called serial reconstructions.

PROCEDURE

The *S*s were 27 introductory psychology students from the Johns Hopkins University.

Experiment I.—The first *S* was presented with a list of letters and asked to memorize its order. Then he was given all the letters—each one typed on a separate card—and was asked to arrange the cards in the correct order. His ordering was typed and presented to the second *S* to memorize and to arrange in order, and this second ordering was typed and presented to the third *S*, and so on.

Eleven successive *S*s were instructed to memorize and reconstruct the order of six lists of scrambled letters. The *S*s were tested individually, one after the other, each one trying to memorize the six orderings of the preceding *S*. The experiment was discontinued after the eleventh ordering because the orderings appeared to have reached a limit of no further approach toward English.

Six sentences ranging from 25 to 33 letters in length (counting spaces as letters) were selected from a second grade reader, and the individual letters in each sentence were scrambled. For instance, HE SAT DOWN might yield SHT EAW NDO.

The first *S* was given 20 sec. to study each list (by complete presentation). Then he was given all the original letters of the list—each letter typed on a separate card—and was asked to arrange them in the correct order. Before he began to arrange them, he was required to write down all he could recall. Because arranging the cards sometimes took him as long as 10 min., as he arranged the cards, he refreshed his memory by referring to these notes if he liked. The notes were discarded, but his ordering of the letters was typed and presented to the second *S* to memorize. The second *S* was given the same cards (after they had been shuffled) and asked to arrange them in order, and this second ordering was typed and presented to the third *S*, and so on.

With each new ordering, the list gradually approached the order of English spelling, so the time given to study a list was gradually reduced. As far as possible, *E* adjusted the time so that *S* made from three to six errors when ordering a list. By the eleventh ordering, the time had been reduced to 12 sec. for a list.

To reduce the serial position effect, the *S*s were told: "Most people remember the first and last portions of a list best. Since you must arrange all the cards, you will do better if you spend proportionately more of your time memorizing the middle of the list."

Experiment II.—For word sequences, Exp. I was repeated with the following changes: (*a*) For materials, six sentences ranging from 18 to 24 words in length were selected from "Storyville Days and Nights" by Louis Armstrong (1954) and the words in each sentence were scrambled. (*b*) The first *S* was given 35 sec. to memorize a list, and by the final ordering this had been reduced to 13 sec. (*c*) Sixteen *S*s (or 16 orderings) were needed before the lists reached a limit of no further approach toward English.

RESULTS

Letter-sequence interferences.—Experiment I clearly showed that *S*s replaced uncommon letter sequences with more common ones. Although

TABLE 1

ORIGINAL LIST AND ELEVENTH ORDERING FOR FIVE LISTS OF SCRAMBLED LETTERS

Original List	Eleventh Ordering
(a) NEHSMFECR DI VARWER HOTIHTHF ER →	MERCHIV REAHER SFDT THON HFWI
(b) HNEWTWY NHETO NT HTKN AOSNDO →	SWYTHE TKNOT ANDNO WEHN TOHN
(c) WFD ALHOD HONIO NOOEK WDE ETTU →	ALH KEETE WOUD HUOTN DOWD FOOI
(d) TEFTLH EAGN HGRE PULA EHT →	TEATHELD TUG NAHL EFH ERG
(e) EHH NC TEHEHM IRATDL IRWCHDE →	CHH CHEEME HUDWAR DIRLT HEIT

they were asked to reconstruct the presented order exactly, with the first four or five reconstructions at least, the lists came closer and closer to English spelling.

Here are several orderings for a list selected at random:

Original	IE TYOUW LHRHES MNHKWEIH CNE ETAR
First	AI THYHLE MUHK WEIO EHRRWN ETCESN
Second	AI HYTWE WENKO HICH ERRHN MELTESN
Sixth	AI WEKNO HICH NERYHUE SWERT METHL
Eleventh	OWI HWE HEUI ACKRE NTHSRN MEYTHEL

The original and the eleventh orderings for the other five lists are given in Table 1.

It may not be obvious that the final orderings in Table 1 resemble English more than the originals, but Fig. 1 shows quantitatively that *Ss* did replace infrequent sequences with ones that more closely resembled English spelling. In Fig. 1, mean frequency per trigram for all six lists is plotted against successive orderings. For every list, all the three-letter sequences were written down. Spaces were not considered characters. (For instance, IE TYOUW LHRH yields TYO, YOU, OUW, LHR, HRH.) The frequency with which each trigram occurred in 141,000[2] English words was determined (Underwood & Schulz, 1960, Total Count); and mean frequency was plotted against successive orderings. For the first four or five orderings at least, it is apparent that *Ss* replaced presented sequences with sequences that more closely resembled English spelling. Tau between frequency and ordering is .36 which is significant beyond the .05 level. Note that 6 of the 11 *Ss* were tested at the horizontal part of the curve; therefore this tau is a very conserva-

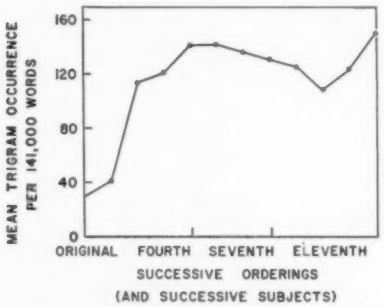

FIG. 1. Mean trigram frequency for six lists of letters arranged in order of successive orderings by *Ss*.

[2] Underwood and Schulz (1960, p. 75) state that their Total Count represents frequencies of occurrence in 1,035,000 words. However, B. Underwood (personal communication) has indicated that this is in error. In Appendix D, the Thorndike-Lorge (1944) count was based on a sample of 2,080 of the 19,440 words in the *The Teacher's Word Book of 30,000 Words*, and so represents a base closer to 106,000 than 1,000,000 words. Thus, the Underwood and Schulz Total Count represents frequencies of occurrence in approximately 141,000 words.

TABLE 2

ORIGINAL LIST AND SIXTEENTH ORDERING FOR FIVE LISTS OF SCRAMBLED SENTENCES

Original List	Sixteenth Ordering
(a) ABOUT WAS GOOD-LOOKING WAY AND TREATING MADE OF THAT A HIM THE QUIET YOUNGSTER NICE HE MANNERS A THEM GIRLS WILD GO WITH	HE WAS A YOUNGSTER NICE QUIET WITH MANNERS GOOD-LOOKING AND A WAY OF TREATING THEM THAT MADE THE GIRLS GO WILD ABOUT HIM
(b) AS BE CHILDHOOD TO LIVED FROM FRIENDS BEEN CONTINUE REAL HAD TRUE AS WE LONG WE AND WOULD WE	WE LIVED AS FRIENDS LONG TRUE FROM CHILDHOOD AND WE WOULD BE REAL TO CONTINUE AS WE HAD BEEN
(c) I BACK AND LONG AND SITTING ALL WORK DAY MULE MY MY IT SHOVELING GET BEHIND TO AWFUL USED PAINS COAL IN HARD WAS	I AND MY ALL DAY LONG SHOVELING COAL IT WAS HARD AWFUL WORK AND SITTING BEHIND MULE USED TO GET PAINS IN MY BACK
(d) GOOD WE BIG AND ME A COOKED IRENE GUMBO SEE OF HIM POT FOR HE AND CAME TO	GOOD CAME ME AND FOR HIM IRENE COOKED A BIG POT OF GUMBO TO SEE HE AND WE
(e) THEM HEART I WORLD OF MY FROM THE TO THE BOTTOM BE REPLACE WILL ABLE NEVER THAT SAY AND	NEVER SAY THAT WORLD AND BE MY HEART ABLE TO REPLACE THE I WILL FROM THE BOTTOM OF THEM

tive estimate of the correlation between frequency and ordering.[3]

Figure 1 shows that after the fifth ordering—after the sequences reached a mean occurrence of about 130 occurrences per 141,000 words—there was little if any further approach toward English spelling. Although the orderings changed somewhat, they remained at about the same level of nonsense. This suggests that the *Ss* deliberately tried to reproduce the same level of nonsense even when they had forgotten the exact sequence.

Word-sequence interferences.—Experiment II repeated the above results for word sequences: the *Ss* replaced uncommon word sequences with more common ones. With each new reconstruction, the arrangement came closer and closer to a sensible English sentence.

[3] These reconstructions were also scored as to bigram frequency per 121,000 words. The curve was almost identical to Fig. 1. It reached a peak at the fourth reconstruction, fell slightly, and began to rise again at the ninth reconstruction.

Several orderings for a list selected for its inherent interest are given:

(Original) STAND DID LINGERIE THEY IN STORYVILLE THE DID AS NOT NEIGHBORHOOD FINE IN DOORWAYS SILK THEIR WEARING GIRLS OUR IN. (First) STAND THEY DID IN STORYVILLE LINGERIE FINE THEIR THE DID NOT IN SILK NEIGHBORHOOD AS DOORWAYS WEARING GIRLS OUR IN. (Second) STAND DID THEY STORYVILLE LINGERIE FINE THEIR IN SILK NEIGHBORHOOD OUR THE NOT GIRLS DID WEARING IN DOORWAYS IN AS. (Fourth) FINE DID THEY STAND IN STORYVILLE LINGERIE IN NEIGHBORHOOD SILK OUR NOT GIRLS THE WEARING DID IN DOORWAYS AS THEIR. (Seventh) AS THEY DID STAND STORYVILLE LINGERIE IN SILK NEIGHBORHOOD IN OUR DOORWAYS THEIR NOT WEARING GIRLS FINE DID IN THE. (Eleventh) AS THEY DID NOT STAND STORYVILLE LINGERIE IN THEIR FINE SILK DOORWAYS IN OUR NEIGHBORHOOD DID THE GIRLS WEARING IN. (Sixteenth) THEY DID NOT STAND IN THEIR FINE SILK STORYVILLE LINGERIE IN THE DOORWAYS WEARING AS OUR GIRLS DID IN NEIGHBORHOOD.

In Table 2, the original and the sixteenth orderings are given for the other five lists. Even if it is not clear from the examples that the sixteenth

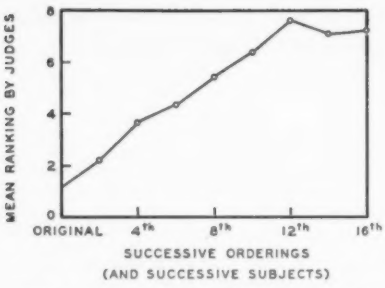

FIG. 2. Mean ranking by judges for six scrambled sentences arranged in order of successive orderings by *Ss*.

orderings resemble English more than the originals, Fig. 2 shows quantitatively that *Ss* did replace unfamiliar word sequences with more familiar ones. In Fig. 2 successive orderings are plotted against mean rank as to sensibleness. For each scrambled sentence, the original and eight even orderings were typed on individual cards, shuffled, and 10 judges were asked to rank them as to their approximation to English.[4] Tau between ordering and mean rank was .89, which is significant beyond .0005.

Figure 2 suggests that the orderings reached a limit of no further approach toward English somewhat before the sixteenth ordering, and the examples show that the sixteenth ordering is still far from a sensible English sentence.

DISCUSSION

The fact that *Ss* replaced uncommon sequences with more common ones is a rather direct demonstration of the extraexperimental interference from language habits, and further discussion would add little of value; however, three procedural details are worth examining. They do not raise serious doubts about the main conclusions, but they may influence the shape of the functions in Fig. 1 and 2.

First we should examine the consequences of *E*'s attempt to hold errors constant by progressively reducing study time as the lists became progressively more simple. For the later *Ss*, this constant-error criterion forced errors on sequences that were almost correct English. Thus it raised the asymptote and reduced the negative acceleration somewhat. The more usual constant time limit produces a sharper negative acceleration: as soon as the list comes close to English, *Ss* are able to memorize it completely and the curve levels off.

Second, each *S* was exposed to and tested upon six successive lists totaling over 130 successive items. On the later lists, he was clearly subject to considerable proactive interference from the earlier sequences and from his own responses on earlier tests. Proactive interference from *S*'s earlier lists should have no effect at all if we assume that there are not extraexperimental interferences from language habits. Interference from lists would cause him to replace one trigram (for instance) with another from an earlier list. Some replacements would be of more common occurrence in English, some of less common occurrence. If the replacements were consistently of more common occurrence, then this would argue for extraexperimental interference from language habits. Proactive interference from earlier tests would lead *S* to replace a trigram with a trigram from his own earlier reconstructions. This would raise mean trigram frequency because the mean trigram frequency of the reconstructions tended to be higher than the lists. Clearly this is a second-stage effect of extraexperimental interference from language habits.

Third, it should be emphasized that the test was not pure reconstruction: it was reconstruction confounded with recall because *Ss* were required to jot down their recall of the list and were allowed to consult these notes during reconstruction. The reconstruction tech-

[4] The judges were all graduate students familiar with approximations to English. Six coefficients of concordance were computed—one for their orderings of each sentence. Mean coefficient of concordance was .88, significant beyond the .001 level. Clearly the judges agreed among themselves.

nique was simply a strategem to prevent the loss of elements that occurs in the usual serial reproduction technique.

In addition to the three procedural points, a marginal finding is worth examining—the finding that long before the reconstructions became sensible English, they reached a limit of no further approach toward English. After this point, although the orderings changed somewhat, they remained at about the same level of nonsense. Apparently *S*s were deliberately trying to reproduce the same level of nonsense.

In fact, several *S*s reconstructing scrambled letters commented that when they were unsure of a sequence, they would be certain that it was not a real word. For if it had been a real word, they said, they were sure they would have remembered it. Therefore in such a situation, they would purposely arrange the letters so that they did not spell a word.

To check the fact of the asymptote, two lists of each kind were reconstructed 10 more times by another group of *S*s. The letter lists dropped from a mean of 128 occurrences per 141,000 words at the eleventh reconstruction to 124 at the twenty-first. In short the fourth reconstruction resembled English spelling about as closely as the twenty-first reconstruction.

Summary

Two extraexperimental sources of interference were examined: (*a*) letter-sequence interferences, which result when a list presented for learning conflicts with English spelling habits, and (*b*) word-sequence interferences, which result when a list presented for learning conflicts with the habits of English syntax. Both interferences were studied by serial reconstructions. The first *S* was exposed to a list and then he tried to reconstruct its order. Each succeeding *S* tried to reconstruct the order given by the preceding *S*. With each successive reconstruction, scrambled lists came closer and closer to sensible English. However, the orderings reached an asymptote after which there was no further approach toward English.

References

Armstrong, L. Storyville days and nights. In, *New world writing*. New York: Mentor, 1954.

Briggs, G. E. Acquisition, extinction, and recovery functions in retroactive inhibition. *J. exp. Psychol.*, 1954, **47**, 285–293.

Postman, L. Extra-experimental interference and the retention of words. *J. exp. Psychol.*, 1961, **61**, 97–110.

Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words*. New York: Teachers College, Columbia University, 1944.

Underwood, B. J., & Postman, L. Extra-experimental sources of interference in forgetting. *Psychol. Rev.*, 1960, **67**, 73–95.

Underwood, B. J., & Schulz, R. W. *Meaningfulness and verbal learning*. Chicago: Lippincott, 1960.

(Received May 29, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 1, 52–57

THE EFFECT OF ORDER OF APPROXIMATION TO THE STATISTICAL STRUCTURE OF ENGLISH ON THE EMISSION OF VERBAL RESPONSES[1]

KURT SALZINGER, STEPHANIE PORTNOY, AND RICHARD S. FELDMAN

Biometrics Research, New York State Department of Mental Hygiene and Columbia University

The variables controlling the emission of verbal behavior may be roughly divided into response-produced stimuli (the speaker's verbal responses surrounding the "controlled" response) and external stimuli (audience, reinforcements, etc.). The present study deals with the effect of response-produced stimuli as specified by Miller and Selfridge's (1950) passages of different orders of approximation to the statistical structure of English.

These passages were originally constructed for the purpose of studying the effect of sequential association upon memory. It was hypothesized that the better memory for higher orders of approximation to English is related to the greater number of words determining each subsequent word. Many investigations have made use of these or similarly constructed passages as independent variables for such factors as memory (Deese & Kaufman, 1957; Marks & Jack, 1952; Miller & Selfridge, 1950; Richardson & Voss, 1960; Sharp, 1958), shadowing of one of two dichotic messages (Moray & Taylor, 1958), "meaningfulness" (Marks & Taylor, 1954), and eye-voice span (Lawson, 1961). Therefore it was thought worthwhile to obtain an exact measure of the amount of contextual dependency in each passage in order to test the assumption that consecutive orders are equally distant from each other. Obviously, if they are not, the shape of any curves employing this assumption in relating the above mentioned factors to approximation to English would be inaccurate.

The technique chosen for evaluating the amount of contextual dependency was the "cloze" procedure (Taylor, 1953, 1954, 1956), which was originated as a measure of readability and which consists of having *Ss* guess the words which have been systematically deleted from a given passage. Thus, application of the cloze procedure to these passages will—in addition to supplying an exact measure of contextual dependency—provide standards for comparing different texts, e.g., one text might be described to be as readable as third-order, another as readable as seventh-order approximation to English.

Finally, this study will provide the opportunity to examine the effect of order of approximation upon the number of guessed words which fall into the same grammatical category

[1] This investigation was carried out while the third author held a predoctoral fellowship from the National Institute of Mental Health, United States Public Health Service. This study was supported in part by Research Grant M1541, in part by Research Grant MY 4758, from the National Institutes of Health. The authors wish to thank J. Zubin for his help and interest in this research. William Reynolds of Rutgers University deserves special thanks for making available a majority of the *Ss*. We appreciate the assistance of Hilda Brody of Columbia University in providing additional *Ss*, and of Robert Keisner and Phyllis Zlotogura for analysis of part of the data. This paper was presented in part at a meeting of the Eastern Psychological Association, Philadelphia, 1961.

of speech as the "correct" word. This will make it possible to evaluate the contribution which the syntax, resulting from these approximations to English, makes to such functions as memory.

Method

Passages.—The 50-word passages of Miller and Selfridge (1950), at each of the eight orders of statistical approximation to English (zero, first, second, third, fourth, fifth, seventh, and Text), were the basic material for this study. Each passage was mimeographed on a separate sheet of paper, each deleted word being indicated by an underlined blank of constant size. The order of presentation of the passages was randomized within each set of eight and varied from *S* to *S*.

Experimental groups.—Ninety-three undergraduate students, with a median age of 18.9 yr., were required to participate in this experiment as part of their regular course work.

Group A_1 (17 *Ss*, 11 male and 6 female) and Group A_2 (16 male *Ss*) were given the passages with every fifth word deleted, starting with the fifth word of each passage and omitting the last word. Groups A_1 and A_2 gave essentially the same results on all measures, and therefore they have been combined into a single group (A_1A_2) in some of the analyses presented below. For all analyses of grammatical category, only Group A_2 is included since it was assumed that further analysis of Group A_1 would have yielded the same results.

Group B (33 male *Ss*) was given the same passages as Groups A_1 and A_2, with every fifth word deleted, starting with the sixth word of each passage and omitting the first word.

Group C (27 male *Ss*) was given the same passages with every seventh word deleted, starting with the seventh word of each passage and omitting the last two words.

Procedure.—The passages were administered to *Ss* in a group. They were told in the instructions read to them that the typescripts contained no punctuation and they themselves would have to decide where a new thought began, that only one word was called for in each blank, and that, when they were not certain, they should guess rather than skip a blank. Thirty minutes were allowed for completion of the task.

Results and Discussion

Proportion of words correctly guessed.—Figure 1 shows that the proportion

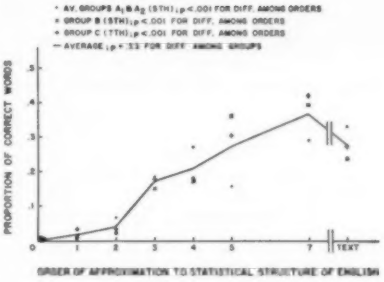

FIG. 1. Proportion of words guessed correctly as a function of order of approximation to English.

of correctly guessed words increased from Orders 0 through 7, with Text approximately equal to Order 5. Friedman analyses of variance (Siegel, 1956), used for this and all subsequent comparisons, showed that the three groups (A_1A_2, B, C) did not differ significantly from one another ($P = .53$) and that, within each group, the increase over orders was statistically significant ($P < .001$ for each group). Therefore, the mean curve in Fig. 1, based on $N = 93$ and on about 40% of the words of each passage, represents a more reliable summary of these data than each curve separately.

This result also makes clear that having 6 words on either side of the blank does not produce more correct guesses than having 4 words on either side of the blank, thus reinforcing the conclusion of Taylor (1954) and MacGinitie (1961) that the words predicted for every fifth blank are independent of each other, and the conclusion of Aborn, Rubenstein, and Sterling (1959) that 11-word sentences result in optimum predictability in comparison to 6-word and 25-word sentences. Apparently *Ss* either do not or cannot make use of a context of more than 5 words on either side of each blank.

The fact that the relationship between order and proportion of correct

TABLE 1

MEAN NUMBER AND PROPORTION OF EXACT CORRECT WORDS AS A FUNCTION OF ORDER OF APPROXIMATION TO THE STATISTICAL STRUCTURE OF ENGLISH

Group	Score	Order of Approximation							Text
		0	1	2	3	4	5	7	
A_1A_2	Number[a]	0	.03	.61	1.64	2.45	1.42	2.61	2.97
	Proportion[b]	0	.003	.068	.182	.272	.158	.290	.330
B	Number	0	.09	.21	1.39	1.58	3.24	3.55	2.15
	Proportion	0	.010	.023	.154	.176	.360	.394	.239
C	Number	0	.29	.29	1.61	1.61	2.72	3.78	2.45
	Proportion	0	.032	.032	.179	.179	.302	.420	.272
Mean Number		0	.14	.37	1.55	1.88	2.46	3.31	2.52
Mean Proportion		0	.016	.041	.172	.209	.273	.368	2.80

[a] Number of correct words per *S*.
[b] Proportion of correct words to total number of words guessed.

words is not linear means that the assumption of equal distances between consecutive statistical orders is not tenable. If we utilize the values in Table 1 as correct estimates of the distances between orders of approximation, then the distance between Orders 2 and 3 is more than five times the distance between Orders 1 and 2. For memory improvement on these same passages[2] (Selfridge, 1949), the corresponding ratio is less than 1.5. Figure 2 shows the memory data plotted in two ways: (*a*) according to the assumption of equal intervals between orders of approximation, and (*b*) with orders spaced according to number of words correctly guessed in the present study. From Orders 3 through 7, the curves are essentially the same. Up to Order 3, however, the intervals between successive orders vary greatly in size, indicating that the equal-intervals assumption is least tenable in this region. One additional fact is that Text gives rise to fewer correct words than Order 7. This indicates that it was misplaced on the order continuum and also suggests that different texts might well appear in different positions on this continuum.

Proportion of different words.—Figure 3 shows a plot of the proportion of the total number of guessed words which were different. With the exception of Text, which is again out of line, it is clear from this graph that increasing order of approximation not only increases the probability of evoking a correct response but also limits the variety of different responses. The difference among the three groups of *S*s approaches significance ($.05 < P < .08$), but the graphs of the separate groups each show the general tendency to decrease over orders. The decrease is significant in each case ($P < .008$ for A_1A_2, $P < .001$ for B, $P = .002$ for C).

Figure 4 shows a plot of the proportion of the total number of different *incorrect* words. With the exception of Text, there is a general decrease in the proportions with increasing

[2] The set of passages used in the present study is available in Miller and Selfridge (1950). However, their memory data are based on averages of two passages at each order of approximation. The percentages of words correctly recalled, for the single set of passages used here, were obtained from Selfridge's (1949) thesis, and correspond to her first set of lists.

order, although the overall drop is only about half of what it is in Fig. 3. The differences among the three groups approach significance ($.05 < P < .08$) as before, but here this is due to the fact that Group C does not show a significant decrease over orders ($P = .20$) while Groups A_1A_2 and B do show significant decreases ($P < .05$ and $P < .04$, respectively). This is the only measure on which one of the groups differed significantly from the others. A suggested explanation is that, with only six blanks for each passage in

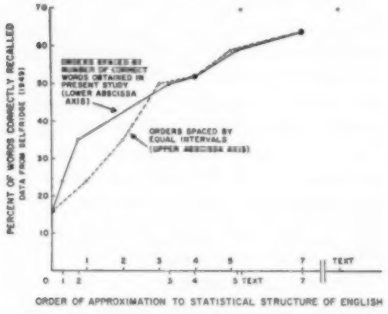
PERCENT OF WORDS CORRECTLY RECALLED
DATA FROM SELFRIDGE (1949)
ORDERS SPACED BY NUMBER OF CORRECT WORDS OBTAINED IN PRESENT STUDY (LOWER ABSCISSA AXIS)
ORDERS SPACED BY EQUAL INTERVALS (UPPER ABSCISSA AXIS)
TEXT
ORDER OF APPROXIMATION TO STATISTICAL STRUCTURE OF ENGLISH

FIG. 2. Percentage of words correctly recalled (Selfridge, 1949) as a function of order of approximation to English, with an equal-interval scale for orders (dashed line), and with the intervals determined by the average number of words guessed correctly in the present study (solid line).

which every seventh word was deleted, the total number of guessed words may not have been large enough to allow significant changes in number of different words from order to order when only incorrect guesses are considered.

Grammatical classification.—The grammatical classification was based on Fries' (1952) system of analysis. The two groups of words called *general* grammatical categories below are lexical words (roughly equivalent to the classical categories of nouns, pronouns, verbs, adjectives, and ad-

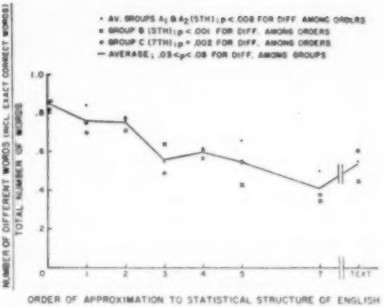
NUMBER OF DIFFERENT WORDS (INCL. EXACT CORRECT WORDS)
TOTAL NUMBER OF WORDS
TEXT
ORDER OF APPROXIMATION TO STATISTICAL STRUCTURE OF ENGLISH

FIG. 3. Number of different words including the correct word (expressed as a proportion of the total number of words guessed) as a function of order of approximation to English.

verbs) and function words (roughly equivalent to articles, conjunctions, prepositions, auxiliary verbs, interjections, and quantity words). The five *specific* grammatical categories are nouns and pronouns considered as one class, verbs, adjectives, adverbs, and function words as already described.

Figure 5 shows that, as order of approximation increases from Orders 0 to 7, an increasing proportion of the words emitted by *S*s belongs to the same specific grammatical category as the deleted words. The increase is significant for each group ($P < .001$ for A_2, B, and C), and the groups do not differ from one another ($P = .24$).

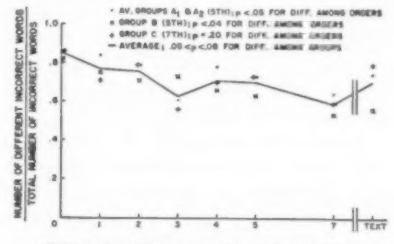
NUMBER OF DIFFERENT INCORRECT WORDS
TOTAL NUMBER OF INCORRECT WORDS
TEXT
ORDER OF APPROXIMATION TO STATISTICAL STRUCTURE OF ENGLISH

FIG. 4. Number of different incorrect words (expressed as a proportion of the total number of incorrect words) as a function of order of approximation to English.

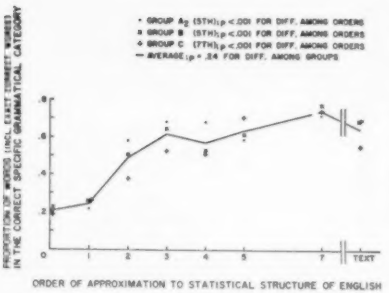

FIG. 5. Proportion of guessed words belonging to the same *specific* grammatical category as the correct word, as a function of order of approximation to English.

This graph again shows performance on Text to be worse than at Order 7 and only slightly better than at Order 5. It should be noted that at Order 0 *S*s perform very close to chance level, i.e., when the emitted words are put into the five specific grammatical categories, about one-fifth fall into the correct category.

Figure 5 differs from Fig. 1 (number of correct words) in that it seems to approach an asymptote at about Order 3 with only a relatively small increase after that, while the number of correct words rises as much after as before Order 3. It is also apparent that the maximum proportion of words in the correct grammatical category is more than twice as large as the proportion of exact correct words, i.e., while *S*s may not react to the meaning they react to the syntax. This type of "knowledge" may well be used as a further way of distinguishing different texts.

Again comparing Fig. 5 and 1, it is seen that the increase from Orders 1 to 2 for words in the correct specific grammatical category accounts for a considerably greater proportion of the overall increase than does the corresponding change for exact correct words. Beyond Order 3, as already stated, correct specific grammatical classifications appear to approach an asymptote while number of exact correct words continues to increase. Therefore, if grammatical classification of guessed words is taken as a measure of amount of syntactical structure, and exact correct words are taken as a measure of amount of meaning, improvement in memory must be attributed primarily to increased syntactical structure between Orders 1 and 2, about equally to syntactical structure and to meaning between Orders 2 and 3, and primarily to meaning beyond Order 3.

If we examine the relationship between order and specific grammatical category, taking into account only those responses which differed from the deleted words, it is found to be nearly identical to that shown in Fig. 5, where correct responses are included. The increases over orders are significant ($P < .001$ for A_2, B, and C) and the three groups do not differ from one another ($P > .15$).

The relationship between order and general grammatical category also remains nearly the same whether or not correct responses are taken into consideration. Again at Order 0 *S*s perform very close to chance level, i.e., when the emitted words are put into the two general grammatical categories of lexical and function words, without further subdivision, about one-half fall into the correct category. In general these relationships are very similar to the analogous ones for specific grammatical categories. Although the increases over orders cover a smaller range when the words are divided into only two categories, they are significant for all three groups (including correct responses: $P < .002$ for A_2, $P < .001$ for B and C; not including correct responses: $P < .02$ for A_2, $P < .001$ for B and C), and the groups do not

differ from one another ($P = .53$ in both cases).

These results are in general agreement with Epstein's (1961) finding that learning of a syntactically structured series of words is superior to that of a random, unstructured series of words, since the two series correspond to Order 1 and Text of the present study. Epstein found in addition that there was no significant difference between his *Ss*' ability to learn syntactically structured nonsense syllable material and their ability to learn syntactically unstructured meaningful material, and this result demonstrates that syntax or meaning alone contributes to learning. This makes more plausible our suggestion that both "meaning" and syntax make contributions to memory for Miller and Selfridge's (1950) passages.

Summary

Ninety-three undergraduate students were required to guess the words that were systematically deleted from a series of passages varying in order of approximation to the statistical structure of English. The *Ss* guessed a greater proportion of words the higher the order of approximation to English. Proportion of words in the correct grammatical category also showed an increase with increasing order of approximation. Proportion of words in the correct grammatical category increased most from Order 0 to Order 3 while proportion of correct words continues to increase as much after Order 3 as before it. Thus analysis of the predicted words into their grammatical categories may well be useful for further differentiation of text materials, since the two types of analysis do not yield the same type of information. Finally, it was pointed out that the assumption of equal intervals between successive orders of approximation is untenable, and the relationship between memory and order of approximation to English can be explained in part by syntactical structure and in part by the "meaning" called for by the context.

References

Aborn, M., Rubenstein, H., & Sterling, T. D. Sources of contextual constraint upon words in sentences. *J. exp. Psychol.*, 1959, **57,** 171–180.

Deese, J., & Kaufman, R. A. Serial effects in recall of unorganized and sequentially organized verbal material. *J. exp. Psychol.*, 1957, **54,** 180–187.

Epstein, W. The influence of syntactical structure on learning. *Amer. J. Psychol.*, 1961, **74,** 80–85.

Fries, C. C. *The structure of English.* New York: Harcourt, Brace, 1952.

Lawson, E. A. A note on the influence of different orders of approximation to the English language upon eye-voice span. *Quart. J. exp. Psychol.*, 1961, **13,** 53–55.

MacGinitie, W. H. Contextual constraint in English prose paragraphs. *J. Psychol.*, 1961, **51,** 121–130.

Marks, M. R., & Jack, O. Verbal context and memory span for meaningful material. *Amer. J. Psychol.*, 1952, **65,** 298–300.

Marks, M. R., & Taylor, W. L. The influence of contextual and goal constraints on the meaningfulness of "automatic sentences." *J. soc. Psychol.*, 1954, **40,** 43–51.

Miller, G. A., & Selfridge, J. A. Verbal context and the recall of meaningful material. *Amer. J. Psychol.*, 1950, **63,** 176–185.

Moray, N., & Taylor, A. The effect of redundancy in shadowing one of two dichotic messages. *Lang. Speech*, 1958, **1,** 102–109.

Richardson, P., & Voss, J. F. Replication report: Verbal context and the recall of meaningful material. *J. exp. Psychol.*, 1960, **60,** 417–418.

Selfridge, J. A. Investigations into the structure of verbal context. Unpublished honor's thesis, Harvard University, 1949.

Sharp, H. C. Effect of contextual constraint upon recall of verbal passages. *Amer. J. Psychol.*, 1958, **71,** 568–572.

Siegel, S. *Nonparametric statistics for the behavioral sciences.* New York: McGraw-Hill, 1956.

Taylor, W. L. Cloze procedure: A new tool for measuring readability. *Journalism Quart.*, 1953, **30,** 415–433.

Taylor, W. L. Application of "cloze" and entropy measures to the study of contextual constraint in samples of continuous prose. Unpublished doctoral dissertation, University of Michigan, 1954.

Taylor, W. L. Recent developments in the use of cloze procedure. *Journalism Quart.*, 1956, **33,** 42–48.

(Received June 5, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 1, 58–61

THE EFFECTS OF REWARD AND KNOWLEDGE OF RESULTS ON THE PERFORMANCE OF A SIMPLE VIGILANCE TASK

RAYMOND R. SIPOWICZ, J. ROGER WARE, AND ROBERT A. BAKER

United States Army Armor Human Research Unit, Fort Knox, Kentucky

Although the effects of reward and knowledge of results on human performance have been intensely studied, the effectiveness of such variables on human monitoring has received relatively little attention. In view of the practical importance of vigilance research and the need to sustain and maintain a high level of detection, this neglect is somewhat surprising. Pollack and Knaff (1958) studied the effects of reward and punishment on monitoring, and reported that punishment was more effective than reward. Moreover, the reward condition proved relatively ineffective in improving performance beyond a level obtained without reward.

Mackworth (1950); McCormack (1959); Loeb and Schmidt (1960); and Weidenfeller, Baker, and Ware (1962) did, however, obtain significant improvement in vigilance performance using knowledge of results (KR). In the last study, not only KR but "false KR" was also found to be effective. Knowledge of results in this experiment was given by means of a bright white light. While it might be assumed that such added stimulation would raise *S*'s activation level and hence increase detection probability, the inclusion of a control group having the light alone (administered in synchrony with the typical vigilance decrement) showed no arousal effects. A similar finding has been reported by Davis, McCourt, and Solomon (1960).

The effectiveness of KR in these studies, and the apparent failure of "reward" in the Pollack and Knaff (1958) study, is somewhat contradictory—especially in view of the fact that feedback or KR can be, and often is, logically regarded as a form of reward or reinforcement. In the Pollack and Knaff study no mention was made of the amount of reward except to say that it was an extra hour of pay. Further, the conditions of reward were such as to make it difficult for more than 1 or 2 *S*s in the group to be rewarded.

If, however, in the typical vigilance study the reward is contingent upon the maintenance of a certain level of vigilance by each *S*, the effects of the reward might be more pronounced than in a situation where a reward is given for each signal detected or on a competitive basis where only the best are rewarded. Moreover, if reward and KR are individually effective, a condition combining the two might be doubly effective.

In order to furnish answers to some of these questions, the effects of KR, reward, and the combination of the two in a typical vigilance task were studied under the conditions in which a monetary reward was individually administered and was contingent upon the maintenance of a high level of performance.

METHOD

Subjects.—Eighty Fort Knox armor trainees, aged 17 to 24 yr., free from visual defects, served as *S*s, Twenty *S*s were arbitrarily assigned to Group C (Control), 20 to Group KR, 20 to Group R (Reward), and 20 to a

combined KR and reward group (Group R + KR).

Apparatus.—The *S*s' task was to detect aperiodic interruptions of a continuous light source over a 3-hr. period. The light source (12 v. dc bulb, .45 ft-c as a point source, operating at 5½ v.) was located at approximately eye level in a flat black plywood box. The schedule of interruptions (signals) was predetermined on the basis of 12 signals per ½ hr. with a total of 72 for the 3-hr. session. The randomized intersignal intervals ranged from 24 to 360 sec. with an average interval of 150 sec. The schedule was presented by means of a film-tape, fed through a Gerbrands variable-interval programer, and a simple timing circuit. The latter determined the duration of the interruption (.03 sec. measured at the timing relay with a Hunter Klockounter), and the former determined the intervals between the interruptions. The signal presentations and *S*s' signal detections were recorded by a 20-pen Esterline-Angus operations recorder.

Procedure.—Each *S*, wearing earphones to reduce ambient noise, monitored the light in an isolated room. For the no-KR condition, *S* was told to press the response button as soon as he saw a signal. Each *S* was further told that the signals could occur at any time, so it was necessary to remain alert and watch the light at all times. To insure that *S* understood the requirements, a practice period of 10 signals (at 1-min. intervals) was given before the watch session began.

For the KR conditions, a 1½-in. pilot lamp was installed at approximately eye level on the right side of the monitoring display. The *S*'s requirements were the same as before; *S* was told that the pilot lamp would light if he missed a signal, thus indicating a signal was missed and that he should be more alert in order to detect subsequent signals. This light, unlike the signal light, was covered by a white filter lens and operated at its maximum voltage—6 v. dc. The light was flashed for a 2-sec. period if a given signal was not detected within 5 sec. after its presentation.

For the reward conditions, each *S* was told he would receive $3.00 if he detected all the signals. In addition, for each signal missed *S* was told that he would lose a portion of the initial $3.00 in geometric progression beginning with 5 cents, for the first miss. By missing six or more signals during the watch session, *S* lost the entire amount. Thus, if *S* missed only one signal in the 3-hr. period he received $2.95; for two misses, he received $2.85; for three, $2.65; for four, $2.25; for five, $1.45, and with six or more misses, nothing. As in the KR condition, a miss was scored if *S* failed to report a signal within 5 sec. after its appearance.

Although the study was designed for 20 *S*s per group, several *S*s fell asleep during the watch session. Since *S*s could not be directly observed, it was arbitrarily decided to eliminate from the study all *S*s who missed 12 or more consecutive signals. This was done on the assumption that *S*s missing 12 or more consecutive signals were either asleep or uncooperative. On this basis, 5 of the original 80 *S*s were rejected—2 from Group C, 1 from Group KR, and 2 from Group R. None of the 20 *S*s, however, were eliminated from Group R + KR. In an attempt to equalize the number of *S*s in each group, additional sessions were held. Due to experimental difficulties, however, data were obtained on only 19 *S*s in Group KR.

Results

The mean percentage of signals missed by each group, by 30-min. periods, is shown in Table 1. On an

TABLE 1

Mean Percentages of Signals Missed for All Groups over Six 30-Min. Periods and over Entire 3-Hr. Session

Group	Successive 30-Min. Periods												Overall	
	7		2		3		4		5		6			
	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*	Mean	*SD*
C	18.2	12.6	21.6	16.5	19.2	13.6	25.0	15.1	27.4	17.4	34.1	27.5	24.3	12.1
KR	7	8.1	10.8	11.1	12.2	11.3	14.4	11.5	14.4	19.1	12.3	17.7	11.8	10.5
R	8.7	11.9	5.9	7.7	5.2	9.5	7.9	8.4	11.6	13.5	10.5	11.2	8.3	7.1
R + KR	4.4	4.9	2.0	3.5	6.2	7.1	5.8	6.7	3.6	4.9	3.6	4.9	4.3	2.4

overall basis, Group C missed 24.3% of the signals; Group KR, 12%; Group R, 8.4%; and Group R+KR, 4.3%. The raw data in terms of number of signals missed by the groups were analyzed to obtain means and variances. A test for homogeneity of variance was performed on these data and resulted in a significant F value. The data were transformed to a log scale and then subjected to an analysis of variance.

This analysis showed the groups were significantly different ($F = 20.92$, $df = 3$, $P < .01$). Using Duncan's (1955) multiple range test, all the experimental groups were found to differ significantly from the control ($P < .01$). The mean difference between Group R + KR and Group KR was also significant at the .01 level. The difference between Group R+KR and Group R was, however, significant at only the .05 level. No significant difference was found between Group R and Group KR.

An inspection of the performance curves also revealed the typical vigilance decrement within the first hour for Group C. This expected decrement, however, was not shown by any of the experimental groups. For the experimental groups, the greatest decrement occurred for Group KR between the first and fourth 30-min. periods. A correlated t test for the decrement, however, showed the difference was nonsignificant.

Although it might be argued that knowledge of results should lead to an increase in the number of false responses, an analysis of the present data showed no significant differences between groups. The *S*s in Group KR averaged 1.05 false responses as compared with 0.84 for Group C, 0.72 for Group R, and 0.25 for Group R+KR.

Discussion

Of most importance was the finding that the combination of reward and knowledge of results led to the detection of better than 95% of the signals—a performance at a "near perfect" level. Further, the extremely small variance of this group may be some indication of the extent to which individual differences in performance can be decreased by the use of such motivational techniques. Classification of *S*s into "good performers" and "bad performers" as was done in other studies (Buckner, Harabedian, & McGrath, 1960; Fraser, 1953) may be adequate in vigilance tasks in which no attempt is made to sustain a high level of motivation. Results from this study, however, suggest the performance of all *S*s in monitoring tasks may be sustained at a high level. Thus, the problem of selection and assignment of *S*s to vigilance tasks on the basis of their initial ability can be eliminated if effective motivational techniques are discovered and properly employed.

Of additional interest is the fact that reward alone proved to be only slightly more effective than KR. It should be noted, however, that in the reward situation *S* had no way of knowing whether or not a singal was missed and, consequently, was unable to adjust effectively his monitoring behavior to meet the requirements of the task.

The relatively high level of performance, as well as the failure to find the usual vigilance decrement for the two reward conditions, focuses considerable attention upon the method used in dispensing the reward. First of all, *S*s did not have to detect all of the signals in order to be rewarded. The *S*, however, was required to detect at least 93% of the signals. Moreover, *S*s were told before the session began that they would be given the reward. Their subsequent performance determined whether or not it would be retained. This technique, it is believed, was particularly effective in sustaining a high level of motivation

over the entire watch period. When specific knowledge of the reward situation was then made available throughout the watch session, *S*'s motivation was further increased and an even higher level of performance was obtained.

In previous studies of KR such as Mackworth's (1950), *S*s were informed whenever they made a false response, when they failed to detect a given signal, and when they correctly detected a signal. It is likely, however, that much of this information is superfluous. Knowledge of results as used in this study gave *S*s information only when they failed to detect a given signal. In terms of the mean number of signals missed, this technique was as effective as those used by Mackworth. Mackworth's experimental group missed 17.3% of the signals presented in a 2-hr. session, whereas *S*s in the present study missed only 12% over a 3-hr. period.

Summary

Four groups of *S*s (20 to a group) monitored aperiodic and brief interruptions of a continuous light source under isolated conditions for a 3-hr. period. The *S*s in Group R were given $3.00 if they detected all signals presented during the watch session, but lost .05, .15, .35, .75, 1.55, or all if they missed one, two, three, four, five, or six signals. Group KR was informed of all signals missed by a bright flash of light. Group R + KR received both KR and reward according to the schedule for Group R. Group C, a control, received neither reward nor KR.

Although all experimental groups were significantly better than the control group, the combination of reward and KR produced the highest level of signal detection. The results are interpreted as indicating that either reward or KR can be effective in maintaining a high level of vigilance. The effectiveness of reward, however, is highly dependent upon the manner in which it is used. The effectiveness of such incentives in improving performance and reducing inter-*S* variability also attenuated the importance previously assigned to individual differences. It is further suggested that such differences are primarily motivational and, as such, are susceptible to experimental manipulation and control.

References

Buckner, D. N., Harabedian, A., & McGrath, J. J. A study of individual differences in vigilance performance. *USA TAGO Hum. Factors Res. Br. tech. res. Rep.*, 1960, No. 2.

Davis, J. M., McCourt, W. F., & Solomon, P. The effect of visual stimulation on hallucinations and other mental experiences during sensory deprivation. *Amer. J. Psychol.*, 1960, **116**, 889–892.

Duncan, D. B. Multiple range and multiple *F* tests. *Biometrics*, 1955, **11**, 1–42.

Fraser, D. C. The relation of an experimental variable to performance in a prolonged visual task. *Quart. J. exp. Psychol.*, 1953, **5**, 31–32.

Loeb, M., & Schmidt, E. A. Influence of time on task and false information on efficiency of responding to pure tones. *USA Med. Res. Lab. Rep.*, 1960, No. 426.

Macworth, N. H. Researches on the measurement of human performance. *Spec. Rep. Ser. Med. Res. Coun., Lond.*, 1950, No. 268, 25–29.

McCormack, P. D. Performance in a vigilance task with and without knowledge of results. *Canad. J. Psychol.*, 1959, **13**, 68–71.

Pollack, I., & Knaff, P. R. Maintenance of alertness by a loud auditory signal. *J. Acoust. Soc. Amer.*, 1958, **30**, 1013–1016.

Weidenfeller, E. W., Baker, R. A., & Ware, J. R. The effects of knowledge of results (true and false) on vigilance performance. *Percept. mot. Skills*, 1962, **14**, 211–215.

(Received June 7, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 1, 62–66

SOME CONDITIONS INFLUENCING THE ACQUISITION AND UTILIZATION OF CUES[1]

LOY S. BRALEY[2]

University of Buffalo

A paper by Bruner, Matter, and Papenek (1955) posed the question as to the variables that control the range of environmental information that an organism acquires during some behavioral sequence. This problem of the "breadth of learning" assumes that an organism is exposed to a wide range of possible cues under most circumstances, but that responses are acquired to only a portion of these with the others remaining as undifferentiated context. While it is clear that the range of cues acquired may be differentially advantageous depending on the task requirements, the focus of their paper and the present study was on the acquisition of cues that were not immediately task-relevant, but which became so at a subsequent time. The above authors demonstrated, as conjectured by Tolman (1948), a reduction in subsequent utilization of initially irrelevant cues as a product of overlearning and higher levels of motivation.

This was but a start on an empirical question. Recent reviews by Easterbrook (1959) and Kausler and Trapp (1960) on cue utilization demonstrate the continued interest in this problem. Easterbrook, concerned with the relation of emotion to behavioral organization took the position that an increase in emotional arousal functions to reduce the utilization of cues. Consequently, degree of behavioral organization depends on task demands for "breadth" of cues. Kausler and Trapp were in substantial agreement although their focus was on the effects of motivation on the utilization of cues in incidental learning. They suggested additional relevant variables of motivation as generalized *D* versus incentive-oriented set, spatial contiguity of relevant and irrelevant cues, and task difficulty.

The present study was designed (*a*) to demonstrate with human *S*s, as Bruner et al. (1955) have done with rats, that exposure to initially irrelevant cues will facilitate later learning when these cues become the relevant discriminanda, (*b*) to test the hypothesis that high scores on a response-inferred motivational variable (Taylor *MA* scale) are inversely related to acquisition and utilization of such cues, and (*c*) to determine whether the presence of the previously correct cue from earlier learning would differentially facilitate or retard the shift to new cues as a function of low and high motive strength, respectively. Certain clinical conceptions of anxiety (Cameron, 1951; Sullivan, 1953) would predict such a perseverative tendency.

METHOD

Subjects.—Eighty extreme scorers on the *MA* scale were selected as *S*s from a sample of 525 undergraduate questionnaires obtained at the University of Buffalo and Springfield

[1] This article is based on a PhD dissertation submitted to the Faculty of the Graduate School of Arts and Sciences of the University of Buffalo. The author is greatly indebted to Ira S. Cohen, under whose guidance this study was conducted, and to Walter Cohen for many valuable suggestions throughout all phases of the investigation.

[2] Now at University of California, Santa Barbara.

College. The upper 20% of this sample had an *MA* score range of 20 to 40, with a mean of 25.3; the lower 20% a range of 0 to 7, with a mean of 4.59. The mean of all scores was 13.77. Twenty-four *S*s were female, 13 of which were high *MA* scale scorers. All *S*s were contacted and asked to participate in such a manner as to avoid any association between the *MA* scale and the experiment proper.

Because of contradictory evidence (Taylor, 1955) on the relationship of IQ to *MA* score, a short form of the Wechsler-Bellevue, Form I (Herring, 1952) was administered. The correlation between IQ and *MA* was .05. Learning rate on the experimental task and IQ revealed a nonsignificant r of $-.22$.

Task and stimulus materials.—The basic task was one of concept learning with simultaneous presentation of a positive and negative instance on each trial. This is similar to the Bruner et al. (1955) procedure and that of Blum and Blum (1949), and Lashley (1942) in their contributions to the continuity controversy. Each instance consisted of two small geometric figures inside a larger geometric figure, drawn on a white 4×6 in. card. The three figures on any card were combinations of circles, triangles, and squares each of which could be red, blue, green, or yellow. An instance then, might be a small red circle, small green triangle, within a large blue square. The stimuli for the three stages of the experimental procedure were as follows:

Stage I: All *S*s learned a concept where the positive cue was two small like-colored figures within a large figure of a different color. Only 12 positives were used as pretesting had indicated rapid learning for this simple concept. Since only two colors appeared on these positives, 12 combinations were possible. This also permitted appropriate combinations of the three types of figures.

The same negative instances were used in all three stages so the possibility of memorization had to be obviated. The 30 negative instances were combinations of 4 values of color, 3 of shape, and 2 of size (small or large figure). In addition, 6 negatives with all figures of the same color were included to avoid concept attainment by a different-colors cue.

Stage II: While all *S*s continued to be reinforced for the Stage I positive cue, half of the *S*s were exposed to 10 trials of irrelevant cues. Five of these consisted of cross-hatching in the large figure and 5 consisted of the substitution of an ʟ shape for one of the small figures. All new irrelevant cues appeared on Stage I positive instances.

Stage III: The positive cues were now either the presence of a cross-hatched large figure or an ʟ shaped small figure. The 30 positive instances were identical to the negatives except for the addition of cross-hatching on half, or substitution of a small ʟ shape on the remainder. No two instances that were the same in original form were opposed during exposure to *S*. Half of the *S*s in Stage III had the Stage I positive cue present as a negative instance on 50% of the trials. For this condition a set of 20 negatives was used consisting of the last 10 Stage I positives and the last 10 of the regular negatives.

Design and procedure.—The experiment was a $2 \times 2 \times 2$ factorial design with 10 *S*s per cell. Conditions were: (*a*) high and low anxiety, (*b*) presence or absence of irrelevant cues in Stage II, and (*c*) presence or absence of Stage I positive cue during Stage III.

The *S*s were randomly assigned to one of the four experimental conditions according to anxiety grouping and were run individually. The IQ measure was obtained at the conclusion of the experimental task.

The *S* faced a plywood screen with two exposure windows at approximately eye level. With the simultaneous presentation of two cards on each trial *S* was to indicate his choice of the correct card by depressing one of two response keys fixed to the table immediately below the windows. A correct choice closed a circuit and reinforcement was delivered immediately by a small white bulb recessed between and above the windows. Position of the correct card was varied according to a prearranged randomly chosen order. The *S*s were instructed that their job was to choose the correct one of two cards and that they would have to begin by guessing. However, the reinforcing stimulus would indicate when they were correct in their choice so that eventually they were to reach a point where they could choose the correct card each time.

In Stage I, trials were continued to a criterion of 10 successive correct responses. Eight *S*s were discarded for failure to meet criterion by Trial 70. Stage II began with completion of the criterion trials and was, to *S*, continuous with Stage I in that there was no interruption in presentation of the stimulus materials. The experimental group was exposed to the series of 10 irrelevant cues. The control *S*s were simply continued for 10 trials beyond criterion on the Stage I cue, giving an equal number of overlearning trials to both groups. Stage III was again continuous with Stage II and where reinforcement could only be achieved by re-

TABLE 1

MEANS AND *SD*s OF TRANSFORMED TRIALS TO CRITERION FOR CONCEPT LEARNING: STAGE III

Anxiety	Positive Cue from Stage I	Irrelevant Cues Present in Stage II		Irrelevant Cues Absent in Stage II	
		Mean	*SD*	Mean	*SD*
High	Present	5.55	2.55	3.17	2.04
	Absent	6.35	2.93	5.68	2.91
Low	Present	5.45	1.51	3.93	1.56
	Absent	8.13	1.16	4.40	2.97

sponding to either of the previously irrelevant cues.

RESULTS

There was no substantial difference in Stage I acquisition trials between control *S*s (Mean = 10.20, *SD* = 11.60) and those exposed to the irrelevant cues in Stage II (Mean = 11.51, *SD* = 12.92). While not shown, it might also be noted that neither experimental nor control *S*s made any errors during the 10 Stage II trials.

Table 1 presents the transformed means and *SD*s of the eight experimental conditions. A Bartlett test revealed nonhomogeneous variances among the conditions, necessitating a square-root transformation to meet the requirements of the analysis of variance.

Table 2 indicates a significant *F* for the Stage II irrelevant cue exposure condition. However, rather than facilitating performance in Stage III, exposure to the irrelevant cues impaired performance. It appears then that the facilitation of subsequent learning through utilization of previously irrelevant cues, as found by Bruner et al. (1955), cannot be demonstrated with human *S*s under these experimental conditions.

It is also clear from Table 2 that differences in the motivational condition (*MA* scale) bear no relationship to performance in the Stage III concept learning. Further, an analysis of variance of Stage I cue learning and *MA* score (within *MS* larger than between *MS* for *df* = 7/72) was not significant.

For the third condition the analysis of variance indicates that continuation of the Stage I positive cue into Stage III facilitated performance independent of *MA* scale level. Difficulty in shifting from a successful (positively reinforced) cue to a new one while the old one is still present was no more characteristic of the high-*MA* *S*s than of the low-*MA* *S*s.

With two cues required for concept attainment in Stage III it is of some interest to determine differential cue difficulty. Error scores were computed for the two cues in the Stage III task and an analysis of variance run on the irrelevant cue exposure condition and the type of cue (cross-hatched ground vs. ⌊ shaped figure). A significant *F* (*F* = 10.62, *df* = 1/156) emerged for the irrelevant cue condition, confirming the analysis by trials, as well as for the type of cue (*F* = 8.88, *df* = 1/156). The cross-hatched cue covering the entire background of the

TABLE 2

ANALYSIS OF VARIANCE OF TRANSFORMED TRIALS TO CRITERION FOR CONCEPT LEARNING: STAGE III

Source	*df*	*MS*	*F*
Anxiety (A)	1	1.69	
Irrelevant cue Exposure (IC)	1	86.15	14.49*
Stage I positive cue present (PC)	1	52.17	8.78*
A × IC	1	6.06	1.02
A × PC	1	.04	
IC × PC	1	.31	
A × IC × PC	1	19.17	3.22
Within	72	5.95	

* $P = .01$.

large figure was clearly an easier cue. With no procedure available for scaling the difficulty of cues it seems reasonable to assume that this was a more novel cue generating higher attention value.

Discussion

The failure of the motivational variable to predict differential acquisition and utilization of cues deserves little comment. Strictly speaking, these results are applicable only to this task, although they raise further questions about the status of the *MA* scale as an independent selection device.

While several studies (Bruner et al., 1955; Jeeves & North, 1956; Peterson & Peterson, 1957; Weiss & Margolius, 1954) have demonstrated positive transfer of context or irrelevant stimuli, the present results suggest additional factors that demand attention. Here, exposure to initially irrelevant cues functioned so as to inhibit rather than facilitate utilization of these cues when subsequently they became the basis for correct discrimination. Berlyne (1958) and Dember (1960) have presented data showing the preference for, and attention value of, stimulus change and novelty. Luria (1957) and Sokolov (1954) have found that orienting reactions are induced by changes in the stimulus field that serve to increase the accessibility of the organism to deal with these changes. These orienting responses are maintained or extinguished as a function of the relevance of new or changed stimuli. Consequently, in the present study, despite the similarity of immediate negative reinforcement at the start of Stage III, the stimuli constituting the positive instances were novel and attention-compelling for the control *Ss* while for the experimental group they represented the familiar. Both the fact that 53% of the control *Ss* reached criterion in the first 10 trials as opposed to only 18% of the experimental group, and the greater ease of learning the cross-hatched ground cue are subject to the same interpretation. It might also be noted that the greater ease in learning the Stage III task where the positive Stage I cue was present as a negative instance fits well with this interpretation. This familiar cue constituted a clearly defined negative instance 50% of the time and thereby served to orient *S* to the other card which was now a positive instance. Had the primary cue been randomly associated with either positive or negative instance in Stage III it would not have had the effect of increasing novelty and, therefore, discriminability of the positive instance.

Two further task variables are probably crucial to the evaluation of the effects of the acquisition of initially irrelevant information. One is that positive transfer of previously irrelevant cues may occur (Babb, 1956; Bruner et al., 1955; Lawrence, 1949, 1950) only where the irrelevant stimuli are not competing in the same sensory field, e.g., a main visual task with irrelevant auditory cues. Different sensory channels would seem to reduce cue interference and the necessity of "adapting out" (Restle, 1955) cues. The second factor is that where the irrelevant cues are randomly present throughout a Stage I task, and then systematically related to the positive cue in Stage II only for the experimental group, then the novelty advantage of these cues for the control group in Stage III should be nullified.

Summary

A three-stage concept-learning task was used to investigate the effects on subsequent cue utilization of, prior exposure to relevant cues, differences in level of motivation (*MA* scale), and the shift of a previously positive cue to the status of a negative instance 50% of the time. The experiment was in the form of a $2 \times 2 \times 2$ factorial design with 10 *Ss* per cell.

The results showed no differences on any aspect of the concept-learning task as a function of level of motivation. The introduction of additional (but irrelevant) cues, after an initial discrimination has been learned, function so as to inhibit rather than facilitate the utilization of these cues when they later become the basis for correct

discrimination. A previously positive cue retained in a later learning task, where its reinforcement value is negative, facilitates the learning of a new discrimination. Results were discussed as suggesting an interpretation in terms of stimulus novelty.

References

BABB, H. Proportional reinforcement of irrelevant stimuli and transfer value. *J. comp. physiol. Psychol.*, 1956, **49,** 586–587.

BERLYNE, D. E. The influence of complexity and change in visual figures on orienting responses. *J. exp. Psychol.*, 1958, **55,** 289–296.

BLUM, R. A., & BLUM, J. S. Factual issues in the continuity controversy. *Psychol. Rev.*, 1949, **56,** 33–50.

BRUNER, J. S., MATTER, J., & PAPENEK, N. L. Breadth of learning as a function of drive level and mechanization. *Psychol. Rev.*, 1955, **62,** 1–10.

CAMERON, N., & MAGARET, A. *Behavior pathology.* Boston: Houghton Mifflin, 1951.

DEMBER, W. N. *Psychology of perception.* New York: Holt, 1960.

EASTERBROOK, J. A. The effect of emotion on cue utilization and the organization of behavior. *Psychol. Rev.*, 1959, **66,** 183–201.

HERRING, F. H. An evaluation of published short forms of the Wechsler-Bellevue scale. *J. consult. Psychol.*, 1952, **16,** 119–123.

JEEVES, A. A., & NORTH, A. J. Irrelevant or partially correlated stimuli in discrimination learning. *J. exp. Psychol.*, 1956, **52,** 90–94.

KAUSLER, D. H., & TRAPP, E. P. Motivation and cue utilization in intentional and incidental learning. *Psychol. Rev.*, 1960, **67,** 373–379.

LASHLEY, K. S. An examination of the "continuity" theory as applied to discrimination. *J. gen. Psychol.*, 1942, **26,** 241–265.

LAWRENCE, D. H. Acquired distinctiveness of cues: I. Transfer between discriminations on the basis of familiarity with the stimulus. *J. exp. Psychol.*, 1949, **39,** 770–784.

LAWRENCE, D. H. Acquired distinctiveness of Cues: II. Selective association in a constant stimulus situation. *J. exp. Psychol.*, 1950, **40,** 175–188.

LURIA, A. R. The role of language in the formation of temporary connections. In B. Simon (Ed.), *Psychology in the Soviet Union.* Stanford: Stanford Univer. Press, 1957. Pp. 115–129.

PETERSON, L. R., & PETERSON, M. J. The role of context stimuli in verbal learning. *J. exp. Psychol.*, 1957, **53,** 102–105.

RESTLE, F. A theory of discrimination learning. *Psychol. Rev.*, 1955, **62,** 11–19.

SOKOLOV, E. N. Higher nervous energy and the problem of perception. In, *Communications to the XIV International Congress of Psychology,* USSR, 1954.

SULLIVAN, H. S. *The interpersonal theory of psychiatry.* New York: Norton, 1953.

TAYLOR, J. A. The Taylor manifest anxiety scale and intelligence. *J. abnorm. soc. Psychol.*, 1955, **51,** 347–354.

TOLMAN, E. C. Cognitive maps in rats and men. *Psychol. Rev.*, 1948, **55,** 189–208.

WEISS, W., & MARGOLIUS, G. The effect of context stimuli on learning and retention. *J. exp. Psychol.*, 1954, **48,** 318–322.

(Received June 12, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 1, 67–76

EFFECT OF STIMULUS CONDITION AND REACTION TIME INFORMATION ON SPATIAL STIMULUS GENERALIZATION [1]

CHARLES Y. NAKAMURA AND JAQUES W. KASWAN

University of California, Los Angeles

Instructed voluntary response (VR) procedures for studying spatial generalization (Brown, Bilodeau, & Baron, 1951) have recently received much attention. But as noted by Sherman and Knopf (1960) little has been done to define the variables that determine error frequency, the measure of stimulus generalization (SG) in this procedure. Although there have been some attempts to generalize formulations derived from classical conditioning situations to the VR situation, there is little agreement between results obtained from the two procedures. One proposed reason for this is that conditioning studies of SG were in large part concerned with primary stimulus generalization, whereas the VR studies must give greater consideration to mediated generalization variables (Gibson, 1959; Mednick & Freedman, 1960).

The concept of stimulus categorization appears to be a useful description of mediational processes involved in VR SG. Used this way categorization involves the assumption that if stimuli to which *S*s are instructed to respond can be discriminated and conceptualized as belonging to a common class, this will facilitate categorization of these stimuli. Where categorization is so facilitated, generalization to other stimuli in the same continuum will be less likely to occur than when categorization is more difficult. The reasoning is similar to Wallach's (1958) discussion of categorization as a determinant in SG-like behavior, where he implied that ease of discrimination of stimuli is directly related to categorizability and inversely related to SG. A recent study by Evans (1961) provides direct evidence that spatial stimulus generalization is a function of whether *S*s can perceive the relative spatial location of the training and test stimuli.

A major goal of the present study is to attempt the identification of some stimulus and response variables necessary to specify the role of categorization in VR spatial SG. A brief description of the experimental situation will assist in making the following discussion of hypotheses more explicit. The stimuli are a horizontal row of 11 lights at eye level on a circular panel presented under three different stimulus conditions. In Cond. 5-6-7, *S*s are instructed to press a key only when Lights 5, 6, or 7 (positive lights) come on. In Cond. 3-6-9, they are to respond only to Lights 3, 6, or 9. In the third condition, essentially a control group, only the center Light 6 is positive.

Because of the spatial contiguity of Lights 5-6-7, *S*s are expected to identify each as belonging to a group of three, thus facilitating categorization of these lights as positive. No such grouping is likely to be discerned in Lights 3-6-9 so that categorization of these lights should be more difficult. Johnsgard (1957), in a study of stimulus-back-

[1] This research was supported by Research Grant G-9589 of the National Science Foundation, and by grants from the Research Committee of the University of California, Los Angeles. The authors wish to acknowledge the contributions of Marion Schulman, who assisted in the collection of the data, and Selma J. Brotsky, who assisted both in the collection and analysis of the data. Parts of this article were presented in a paper read at the 1961 American Psychological Association convention in New York City.

ground contrast, presented a similar explanation of SG effects.

This arrangement of training and test stimuli also permits the comparison of error frequencies predicted by the categorization hypothesis and those predicted by Hull's gradient summation hypothesis based on the exponential summation of overlapping gradients of generalized habit strength. Bilodeau, Brown, and Meryman (1956) reported support for the latter hypothesis. Kalish and Guttman (1957), on the other hand, found only partial support for the Hullian hypothesis both in their review of the Bilodeau et al. (1956) results and in their own research. Under the conditions of the present experiment, the gradient summation hypothesis would predict approximately the same number of errors to Lights 4 and 8 in both Cond. 5-6-7 and 3-6-9. (Lights 4 and 8 were selected for specific comparison of errors because in both stimulus conditions they are located ideally in the spatial continuum where there are at each negative light three overlapping gradients generated from the training lights.) If, however, spatial contiguity facilitates categorization, fewer errors to these two lights would be expected in 5-6-7 than in 3-6-9. Moreover, in accord with the categorization hypothesis that the three positive lights in 5-6-7 are conceptualized as belonging to a single class, both the total number of errors and the error gradients in 5-6-7 are expected to be similar to those obtained in Stimulus Condition 6 where only the single light Number 6 is positive. Such a prediction would not be made from the gradient summation hypothesis.

Specific predictions of response latencies also follow from the categorization hypothesis. On the assumption that reaction time (RT) generally increases as a function of stimulus complexity (Flores, 1956; Grebb, 1954) and that categorization is presumably a determinant in accuracy of perception (Bruner, Goodnow, & Austin, 1956, p. 9), it would be expected that accurate response latencies should be shorter in 5-6-7 than in 3-6-9.

The above predictions of errors and latencies generated from the stimulus categorization hypothesis produce, as logical consequents, some specific predictions of error-latency relations. If latencies reflect difficulty of categorization of positive lights, *S*s who do not take the required time for accurate categorizing should make more errors than those who take adequate time. Accordingly, if *S*s can be induced to reduce response latencies they should make more errors. Since there is some evidence that knowledge of performance may reduce RT (Bilodeau & Bilodeau, 1961, p. 250), giving *S*s information about their speed of reaction should reduce latency. Insofar as 5-6-7 latencies are expected to be close to simple RTs, further reductions in RT or increase in errors should be limited. In 3-6-9, however, greater absolute reductions in latencies are possible. Accordingly, greater amount of speed information should be associated with larger error differences and smaller latency differences between the two stimulus conditions, and these should be attributable to the speed information effect in Cond. 3-6-9.

Method

Apparatus.—The apparatus was a modification of that used by Brown et al. (1951). The main component was a 180° semicircular panel on which was mounted a horizontal row of 11 lights separated by intervals of 4° of visual angle, at eye level. The essential modification was the use of 4° of visual angle with the lights at a distance of 30 in. from *S*'s eyes instead of the 8° intervals at a distance of 60 in. employed by Brown et al. The lights, identified by numbering 1 through 11 from left to right, were standard 6.3-v., .15-amp. miniature bayonet indicator lamps encased in milk-glass covers. A similar orange colored light located directly above the central light, Light 6, served as both fixation point and ready signal. The panel was 3 ft. high, painted flat black, covered over the top with black cloth, and placed in a semidark room to minimize extraneous visual cues. The *S* responded by pressing a telegraph key placed at his right hand (or left, if left handed). Trial intervals were regulated by an electric

motor. The *E*, seated behind the panel and not visible to *S*, selected the stimulus light, randomly varied the foreperiod between 2, 3, and 4 sec. on a Hunter electric interval timer, and read the RTs to the nearest .01 sec. from a Hunter electric clock.

Procedure.—The basic 2 × 3 factorial design involved six sets of *Ss* under two stimulus conditions and three RT information groups. The groups under Cond. 5-6-7 were instructed to respond to the lighting of either Lights 5, 6, or 7 and those in Cond. 3-6-9 to either 3, 6, or 9. All the groups were instructed not to respond when any of the other lights were lit.

Within Cond. 5-6-7 and 3-6-9, *Ss* in the total information (TI) group were given information of speed of response after every trial and the partial information (PI) group after every fifth trial. The *Ss* were told their performance was "good" (short latency), "fair," or "poor" (long latency) in equal proportion and *E* attempted to make this correspond with *Ss*' actual variations in latencies. For the no information (NI) group, speed and accuracy were stressed equally and instructions to this effect were repeated after every 25 trials. All *Ss* were told when they made an error.

After the sequence of ready signal, stimulus light, and *S*'s response was demonstrated, *S* was given a series of 20 training trials distributed among the three positive lights in a predetermined random sequence. Trial intervals were 12 sec., beginning with the onset of the ready light. Following the training trials, *S* was told that now any of the lights might go on but he was to respond only to the positive lights (3, 6, 7, or 3, 6, 9, depending on the stimulus condition) and not to any other lights. The test series involved 5 presentations of each of the negative lights, along with 40 booster trials of each of the three positive lights distributed in a predetermined random order.

In addition to Cond. 5-6-7 and 3-6-9 in the basic design, a third stimulus condition (Cond. 6) was included where *Ss* were instructed to respond only to Light 6. This group was given total RT information similar to the other TI groups. The number of booster trials given Light 6 during the test series was equal to the sum of booster trials to the three positive lights in each of the other two stimulus conditions.

Subjects.—The *Ss* were 130 students, both men and women, from a course in introductory psychology. Within each of Cond. 5-6-7 and 3-6-9, 20 *Ss* were randomly assigned to each PI and TI group and 15 to each NI speed information group. There were 20 *Ss* in Cond. 6.

Results

Errors.—The mean percentages of errors to each of the negative lights are plotted for the different speed information conditions in Fig. 1 for 5-6-7 and Fig. 2 for 3-6-9. The analysis of variance of total errors over all information groups showed that the stimulus condition effect was significant ($F = 38.19$, $df = 1/104$, $P < .001$) with fewer total errors in 5-6-7 than in 3-6-9. The stimulus effect was also significant for errors to Lights 4 and 8 ($F = 14.91$, $df = 1/104$, $P < .001$), with the fewer errors in 5-6-7. This is in contrast to the nearly identical averages of the summation hypothesis predictions to these two lights of 54% and 52% errors

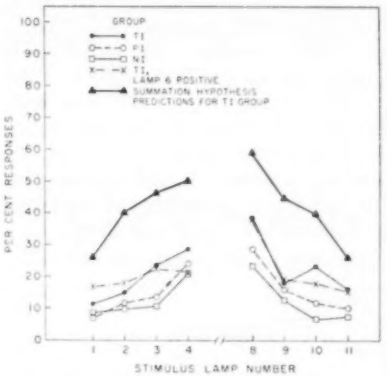

FIG. 1. Mean percentage of errors in Cond. 5-6-7 to each negative (test) lamp, and percentages predicted by the summation hypothesis for the TI group. (The latter were obtained by erecting around each of Lights 5, 6, and 7 the gradient obtained when only Light 6 was positive, and exponentially summating overlapping points. The two wings of the single Light 6, also shown, are displaced one spatial unit away from center so that there is correspondence of percentage of errors to each test light, for all gradients, on the dimension of distance from the nearest positive light.)

expected for 5-6-7 and 3-6-9, respectively. (Lights 4 and 8 frequencies were combined in the analysis of variance since a chi square test of the difference between errors to these lights was not significant—$\chi^2 = .90$, $df = 2$, $P > .10$.) These results are consistent with the stimulus grouping hypothesis that the spatial contiguity of positive lights facilitated categorization and led to the reduction of errors. Errors are clearly not attributable to a failure to discriminate lights as positive since *S*s failed to respond appropriately to positive lights in only three trials. The hypothetical summation gradients in Fig. 1 and 2 were plotted by erecting around each of Lights 5, 6, and 7 and around Lights 3, 6, and 9 the gradient obtained when only Light 6 was positive. The values of the overlapping points, at each negative light, were summated according to Hull's formula for exponential summation of gradients (Hull, 1943, p. 200). The value 100 was substituted for M in the formula where M is the physiological limit of the learning process. In this experiment, the maximum percentage of responses to the positive lights was 100%.

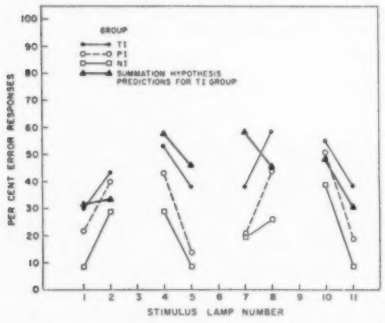

FIG. 2. Mean percentage of errors in Cond. 3-6-9 to each negative (test) lamp, and percentages predicted by the summation hypothesis for the TI group. (The latter were computed as described for Fig. 1.)

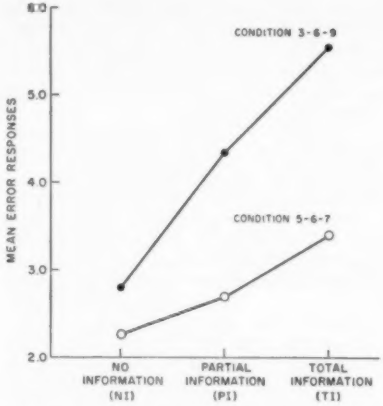

FIG. 3. Mean number of errors to Lights 4 and 8 in two stimulus conditions and three speed information groups.

Figure 1 shows that the error gradient of the TI group in Cond. 5-6-7 closely approximated the gradient obtained in Cond. 6, also under TI. The mean difference between these sets of points was not significant ($F = 1.57$, $df = 1/38$), indicating that the number of stimuli to be categorized did not affect generalization. Moreover, the large discrepancy between summation hypothesis predictions and obtained error frequencies for the TI group in Cond. 5-6-7 indicates that summation does not account for the results obtained by this VR procedure. The mean difference between these two sets of points was significant ($F = 23.81$, $df = 1/8$, $P < .01$).[2] However, Fig. 2 suggests

[2] In order to obtain scores for a variance estimate of the hypothesized gradient, necessary to the analyses of differences between distributions, the summation procedure was carried out separately for Cond. 5-6-7 and 3-6-9 by using the error scores, to the relevant lights, made by each *S* in Cond. 6. A statistical problem posed because several *S*s in Cond. 6 had zero error scores could not be resolved by a transformation since the subsequent scores were to be exponentially summated. This was overcome by using

that in Cond. 3-6-9 the results obtained for the TI group approximated fairly closely the predictions of the summation hypothesis. The nonsignificant F of 2.39 ($df = 1/8$) for the difference between these curves was consistent with this observation.

The curves in Fig. 3 plot mean errors to Lights 4 and 8 as a function of speed information for Cond. 5-6-7 and 3-6-9 and show that greater amounts of speed information were associated with higher occurrence of errors ($F=7.39$, $df=2/104$, $P<.01$). The information effect was also found to be significant for total errors to all negative lights ($F = 12.15, df = 2/104$, $P< .001$). A further t test analysis of errors to Lights 4 and 8 at each information level showed that the stimulus condition effect occurred only where information was given ($t = 2.56$, $df = 38$, $P < .05$, for the difference between PI groups and $t = 3.24$, $df = 38$, $P < .01$, for the TI groups). The corresponding difference between the NI groups was not significant.

Latencies.—All the latency computations were based on the reciprocals of the latency scores (Edwards, 1960, p. 131) and the weight of the statistical analyses was almost wholly on the latency scores to positive lights. The latter was necessary since the latencies for errors were not suited to the usual statistical tests of differences between means since mean latencies for *S*s were computed from unequal numbers of scores because individual *S*s made different numbers of errors to different lights.

Figure 4 shows the marked effect of stimulus condition on latencies to positive lights at all levels of RT informa-

the mean scores of *S*s randomly combined into groups of 4, at the cost of reduction in *df*.

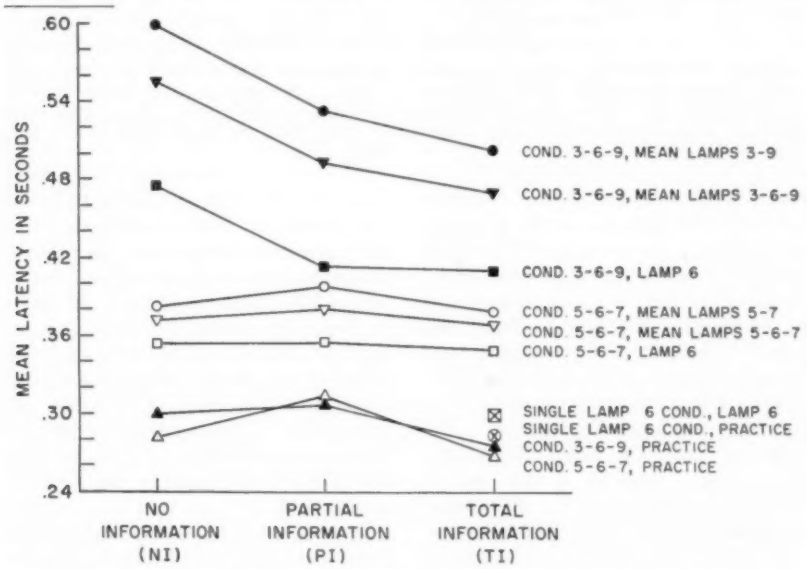

FIG. 4. Mean latencies to positive lights under different stimulus conditions within different speed information groups.

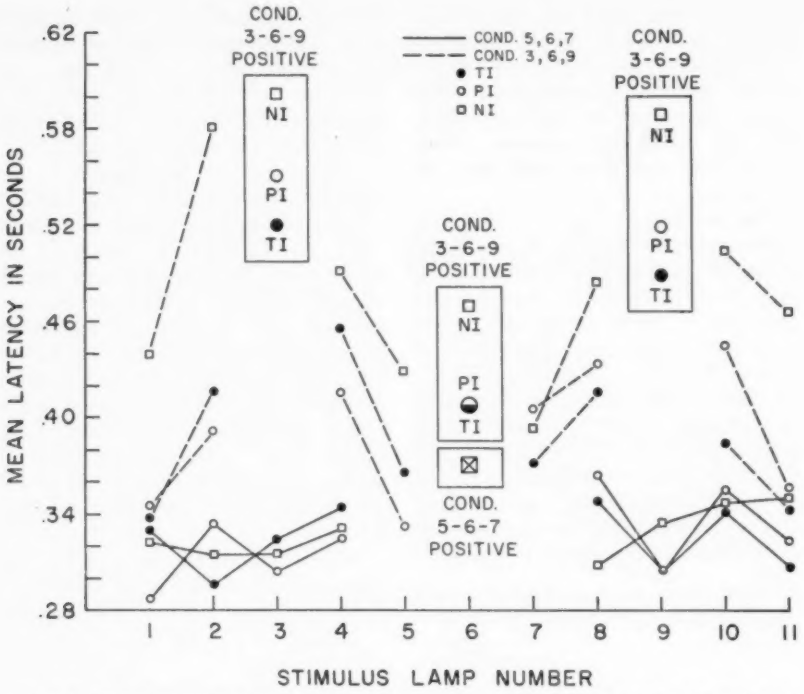

FIG. 5. Mean latencies of error responses to negative (test) lights in Cond. 5-6-7 and 3-6-9. (Mean latencies to positive lights are shown in boxes in the appropriate positions between negative lights. The lowest box above Light 6 shows the combined mean latency of the three information groups to positive lights under Cond. 5-6-7 since the three group means were nearly identical.)

tion. The F of 148.55 ($df = 1/104$) for the test between mean latencies to Lights 3, 6, and 9 and to Lights 5, 6, and 7 of the test trials was significant ($P < .001$). Comparisons between Cond. 5-6-7 and 3-6-9 at each level of information yielded t test values significant beyond the .001 level. The t's were 8.94, 6.41, and 6.49 for comparisons between the pairs of NI, PI, and TI groups, respectively. Figure 4 also shows that the magnitude of the differences between stimulus conditions decreased with the giving of greater amounts of speed information and indicates that the locus of this effect was in Cond. 3-6-9. The apparent differential effect of RT information on latency was confirmed by the significant Stimulus × Information interaction ($F = 4.47$, $df = 2/104$, $P < .02$). Analyses by t tests of the differences between NI vs. SI and NI vs. TI within Cond. 3-6-9 yielded t's of 3.05 ($P < .01$) and 3.84 ($P < .01$), respectively; whereas, none of the t's between information levels within Cond. 5-6-7 reached the .05 level of significance.

It is seen in Fig. 4 that latency to Light 6 in Cond. 3-6-9 was shorter than the mean latency to Lights 3 and 9 ($F = 46.84$, $df = 1/104$, $P < .001$)

and similarly for the corresponding difference between Light 6 and the mean for Lights 5 and 7 in Cond. 5-6-7 ($F = 12.99$, $df = 1/104$, $P < .001$). It is probable that the shorter latency for Light 6 is attributable, in part, to the signal light above it which served as a discriminative cue. However, it is unlikely that this alone could account for the main effects since the absolute mean latency to Light 6 differed among stimulus conditions. It was longest under Cond. 3-6-9, shorter in Cond. 5-6-7, and shortest in Cond. 6, as shown in Fig. 4.

Comparison of the mean of the three mean latencies to positive lights for the TI group in Cond. 5-6-7, with the mean latency to Light 6 for Cond. 6 yielded an F of 18.73 ($df = 1/38$, $P < .001$). This finding contradicted the assumption that latencies to the grouped stimuli should approximate those obtained when a single light was positive. Hence, though the error gradients did not differ for these two stimulus conditions, as noted in Fig. 1, the data suggest that categorizing three contiguous stimuli as a class is in fact more difficult than distinguishing a single positive stimulus in a continuum of stimuli.

Figure 5 plots error response latencies to the negative lights for the three RT information groups within each of Cond. 5-6-7 and 3-6-9. The upper part of Fig. 5 shows that in 3-6-9, latencies for PI and TI groups tend to be shorter than those for the NI group. In contrast, (lower part of Fig. 5) there were no substantial differences in latency gradients or in latency to individual lights as a function of speed information in Cond. 5-6-7, and error latencies were clearly much shorter than those in Cond. 3-6-9 with practically no overlap. Also, error latencies were consistently much shorter than latencies to adjoining positive lights for all three information groups within Cond. 3-6-9 and the differences were slight but in the same direction for Cond. 5-6-7.

Error-latency interaction. — Comparison of latencies to positive lights in Fig. 4 with errors for Cond. 5-6-7 and 3-6-9 in Fig. 1 and Fig. 2, respectively, shows that although latency differences between the two conditions decreased with greater RT information, differences in error frequency generally increased, largely because of the greater effect of speed information in 3-6-9 than in 5-6-7. Wherever RT information reduced latencies to positive lights, as in 3-6-9, error frequency increased. Where RT information had no substantial effect on such latencies, as in 5-6-7, there was no significant difference in error frequency between the information groups within that stimulus condition.

Further evidence of the interaction of errors and latencies was observed in that all except 1 of the 15 *S*s in the NI group of Cond. 3-6-9 obtained mean latencies to positive lights of .5 sec. or greater, whereas, 19 of the 40 *S*s given RT information (PI and TI groups) in Cond. 3-6-9 had mean latencies shorter than .5 sec. Since the NI group obtained significantly fewer errors than the PI and TI groups in 3-6-9, the .5-sec. latency may be conceived as an optimum time for categorization of stimulus lights into positive and negative classes. Accordingly, *S*s responding faster than that should make more errors than *S*s with latencies of .5 sec. or greater. When this comparison was made within TI and PI groups of 3-6-9, an F of 19.85 ($df = 1/38$, $P < .001$) was obtained. Moreover, comparison of error frequency of *S*s above and below the median latency of .58 sec. within

the NI group in Cond. 3-6-9 also yielded a significant result ($t=2.26$, $df = 28$, $P < .05$) for further indication of the relation of shorter latency with greater errors. Similarly for Cond. 5-6-7, though neither error frequency nor latency had differed as a function of RT information, frequency of errors between *S*s above and below the median latency of .4 sec., for the three RT information groups combined, was significantly different ($F=4.19$, $df=1/53$, $P<.05$).

A third source of evidence for error-latency interaction was the negative Pearson product-moment *r*'s between frequency and latency of errors. For Cond. 5-6-7, the *r*'s for each group were NI = −.51, PI = −.41, and TI = −.50. Under Cond. 3-6-9, they were NI = −.48, PI = −.81, and TI = −.71. All except the *r* of −.41 were significant at the .05 level or better. As expected, errors and latencies were most substantially correlated when RT information was given in Cond. 3-6-9, where stimulus categorization was presumably most difficult. However, the only significant difference ($P < .05$) between correlations between Cond. 5-6-7 and 3-6-9 within any information level was that between the −.81 and −.41 for the PI groups.

When percentages of errors in Fig. 1 and 2 are compared with error latencies in Fig. 5, the direction of changes in magnitude from light to light are notably parallel. Though this is particularly true of Cond. 3-6-9, it is also seen in Cond. 5-6-7; where decreases in error are greatest, as from Lights 4 to 3 and from Lights 8 to 9, error latencies show a corresponding decrease in five of the six instances. While latency gradients beyond these points are somewhat irregular, variations are within a narrow range suggesting an essentially horizontal distribution consistent with the flat error frequency gradients. Thus the results suggest that error frequency and error latency *gradients* are *positively* related.

Discussion

The assumption that stimulus conditions facilitate the categorization of lights as positive or negative was consistently supported. It seems, however, that the stimulus contiguity hypothesis does not entirely account for this facilitation. The persistent sharp drop in errors and latencies to lights adjoining Light 6 in Cond. 3-6-9 and the shorter latency to Light 6, compared to other positive lights, suggest that the central position of this light and (or) the signal light located above it served as discriminative cues which facilitated categorization of the adjoining lights as negative stimuli. It is possible that the central location of the positive lights and the signal light also played a role in Cond. 5-6-7, although the first negative light is two lights removed from the signal light. Occurrence of such effects had not been expected since Bilodeau et al. (1956), using a similar apparatus, found that error gradients and frequencies to single positive lights located 16° of visual angle from center were virtually identical with those obtained to a central positive light. Though this finding clouds the interpretation of results in terms of a stimulus pattern effect, the influence of the latter is apparent in the fact that latencies to Light 6 itself decrease from Cond. 3-6-9 to 5-6-7 to Cond. 6, suggesting an effect of stimulus arrangement on speed of categorization.

Specification of stimulus arrangement and stimulus position as determinants of SG appear of special interest because this involves the problem of specifying the role of stimulus units in SG. In their review, Mednick and Freedman (1960) indicate that most SG studies follow an empirical formulation which assumes that SG is a decreasing function of increasing physical difference between

test and training stimuli. Hull (1943, p. 198) explicitly recognized the untenability of postulating such arbitrary simple physical stimulus units and proposed the jnd as the stimulus unit for establishing SG gradients. The present results re-emphasize that equal physical distance units from different positive lights do not correspond to equal decrements in SG in the VR paradigm. This was shown most explicitly by the comparison of the error frequencies obtained in Cond. 5-6-7 with the higher frequencies of Cond. 3-6-9, and between the obtained curves and the summation hypothesis predictions in each of these conditions. It was further demonstrated by the fact that error frequencies to lights adjoining Light 6 in Cond. 3-6-9 were much lower than errors to lights surrounding Light 3 and Light 9, although the difference in physical distance was identical. The results indicate the advantage of a stimulus unit or scale that could take into account the specification of the relation among stimuli. The proposal of an ordinal scale of stimulus values by Mednick and Freedman (1960) is an attempt in this direction. However, their approach would predict sharper gradients over the negative lights in Cond. 5-6-7 than obtained in the present study. Since Lights 4 and 8 are separated from the training stimuli by only one ordinal unit, Mednick and Freedman would expect higher percentages of errors on them, relative to the more peripheral lights, because their formulation does not take into account the effect of contiguity of the training lights.

The overall results of this study show a close relation between error and latency. The negative error-latency relation wherein *Ss* who took less time before responding to training stimuli generally had greater error frequencies indicates that errors were due largely to insufficient time taken (for accurate categorization) before responding. Another aspect of the close relation between categorization and latency is observable in the positive relation between error frequency and error latency which is also found in most other VR SG studies (Gibson, 1939; Mednick, 1958; Rosenbaum, 1953).

The manifold role of latency in the results of this study make it reasonable to suggest that latency may serve to specify both stimulus and response anchors of the term "categorization" used here to describe the verbal and other mediated processes which are undoubtedly important determinants of SG in any VR paradigm. Stimulus referents of categorization can be specified by noting that the time taken for responses varied with the arrangement or position of lights, as in the different latencies for Cond. 5-6-7 and 3-6-9. Response referents of categorization can be specified by noting that latency was a major determinant of error frequency under given stimulus arrangement conditions, as in the effect of speed information in Cond. 3-6-9. Thus, difficulty or ease of categorization is measurable in terms of latency required to make a correct response, under specified conditions such as, in this study, relative emphasis on speed induced by providing reaction time information.

Summary

This study examined the effects of stimulus arrangement and experimentally induced differences in response latency on performance in a spatial stimulus generalization situation involving voluntary responses. In Cond. 5-6-7, *Ss* were instructed to respond to any one of three centrally and contiguously located lights (Lights 5, 6, and 7) of 11 lights in a row, but to inhibit responses to all other lights. In Cond. 3-6-9, the positive lights were noncontiguous (Lights 3, 6, and 9). The results showed that error frequency and latency were smaller in Cond. 5-6-7 than in Cond. 3-6-9. It was also found that giving *Ss* information on speed of responses reduced latencies and increased errors in Cond. 3-6-9 but had practically no effect in Cond. 5-6-7. These findings were consistent with predictions of stimulus generalization generated from a stimulus categorization hypothesis. Predictions of error frequency by this hypothesis were compared with those made from Hull's gradient summation hypothesis of stimulus generalization. The latter failed to predict any of the obtained points on the gradient for Cond. 5-6-7, but did approximate

most of the points obtained in Cond. 3-6-9, under conditions where *Ss* were given maximum speed information. The close and consistent relationship found between error frequency and latency was interpreted as indicating that latency can be used to specify both stimulus and response anchors of the mediational process in spatial VR SG.

REFERENCES

BILODEAU, E. A., & BILODEAU, I. McD. Motor-skills learning. *Annu. Rev. Psychol.*, 1961, **12**, 243–280.

BILODEAU, E. A., BROWN, J. S., & MERYMAN, J. J. The summation of generalized reactive tendencies. *J. exp. Psychol.*, 1956, **51**, 293–298.

BROWN, J. S., BILODEAU, E. A., & BARON, M. R. Bidirectional gradients in the strength of a generalized voluntary response to stimuli on a visual-spatial dimension. *J. exp. Psychol.*, 1951, **41**, 52–61.

BRUNER, J. S., GOODNOW, J. J., & AUSTIN, J. A. *A study of thinking.* New York: Wiley, 1956.

EDWARDS, A. L. *Experimental design in psychological research.* New York: Rinehart, 1960.

EVANS, W. O. Two factors affecting stimulus generalization on a spatial dimension. *J. exp. Psychol.*, 1961, **61**, 142–149.

FLORES, I. The effect of organization upon complex reaction time. *J. Psychol.*, 1956, **41**, 301–313.

GIBSON, E. J. Sensory generalization with voluntary reactions. *J. exp. Psychol.*, 1939, **24**, 237–253.

GIBSON, E. J. A re-examination of generalization. *Psychol. Rev.*, 1959, **66**, 340–342.

GREBB, L. W. The effect of stimulus complexity on discriminative responses. *J. exp. Psychol.*, 1954, **48**, 289–297.

HULL, C. L. *Principles of behavior.* New York: Appleton-Century, 1943.

JOHNSGARD, K. W. The role of contrast in stimulus intensity dynamism (*V*). *J. exp. Psychol.*, 1957, **53**, 173–179.

KALISH, H. I., & GUTTMAN, N. Stimulus generalization after equal training on two stimuli. *J. exp. Psychol.*, 1957, **53**, 139–144.

MEDNICK, S. A. Gradients of latency in a generalized voluntary response. *Amer. J. Psychol.*, 1958, **71**, 752–755.

MEDNICK, S. A., & FREEDMAN, J. L. Stimulus generalization. *Psychol. Bull.*, 1960, **57**, 169–200.

ROSENBAUM, G. Stimulus generalization as a function of level of experimentally induced anxiety. *J. exp. Psychol.*, 1953, **45**, 35–43.

SHERMAN, J., & KNOPF, I. J. Changes in the gradient of stimulus generalization as a function of two procedural variations. *Psychol. Rep.*, 1960, **7**, 253–258.

WALLACH, M. A. On psychological similarity. *Psychol. Rev.*, 1958, **65**, 103–116.

(Received June 27, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 1, 77–80

STIMULUS GENERALIZATION AS A FUNCTION OF THE FRAME OF REFERENCE[1]

DAVID R. THOMAS AND CHARLES G. JONES

Kent State University

In a study reported by Philip (1952) it was shown that the location of the original stimulus within a generalization test series modifies the shape of the obtained gradient. Philip required his *S*s to rank cards containing varying proportions of green and blue dots along a greenness-blueness scale. The frequency with which a given rank, say Number 3, was subjectively attributed to the different cards constituted a gradient of generalization for that value. A separate gradient was thus generated for each rank employed.

Philip systematically varied the length of the generalization test series. With the shortest series employed (six values) there was a tendency for judgments to accumulate near the center of the scale, the "central tendency effect" (Hollingworth, 1909). This effect was reflected in asymmetrical gradients around stimulus values which were noncentrally placed in the series of generalization test stimuli.

Because of the unusual nature of Philip's (1952) procedure, the significance of his finding with regard to generalization as studied by other methods may be questioned. The purpose of the present study was to assess the generality of Philip's finding, using a method for obtaining generalization gradients developed by Kalish (1958). The Kalish procedure is more typical of generalization studies in that *S*s are first exposed to a single stimulus value and are subsequently tested for their ability to select the original from a randomly presented series of stimuli. This is in contrast to the Philip procedure in which absolute judgments of predominating color are made without previous exposure to some standard stimulus value. It was reasoned that if the "central tendency effect" were shown to distort measures involving retention as well as absolute judgment, its relevance for studies of generalization would be more convincingly demonstrated.

[1] This research was supported in part by National Institutes of Health Grant M-4203.

METHOD

Subjects.—The *S*s were 50 undergraduate men taken from introductory psychology courses at Kent State University. All *S*s had normal color vision, as determined with the Dvorine (1944) color perception test.

Apparatus.—The study employed a Skinner-type key pecking apparatus, modified for use with human *S*s. The box was approximately 15 in. long, 11 in. high, and 14 in. wide and was painted flat black. The front wall of the box was made of transparent Plexiglas so that *S*s could clearly view the pecking key. The *S* sat in a chair approximately 2½ ft. from the key, which was a circular plastic disc ¾ in. in diameter. Illumination was provided by a Cambridge Thermionic Corporation monochromator, Model B, Series 1066, equipped with an Olympus, Model 201250 6-v., 5-amp. light source. The patch of color on the key was approximately 4 mL. in luminance. The only other light in the room was a 7.5-w. "night light" on *E*'s side of a black cloth screen separating *S* from *E*.

A telegraph key was used to measure *S*'s responses. The telegraph key was placed on the table next to the Skinner box, within easy reach of *S*'s right hand. It was wired so that its release would illuminate a signal light on *E*'s side of the screen, thus signifying a response. The box was equipped with an electrically operated shutter which interrupted the monochromator beam when *E* threw a switch.

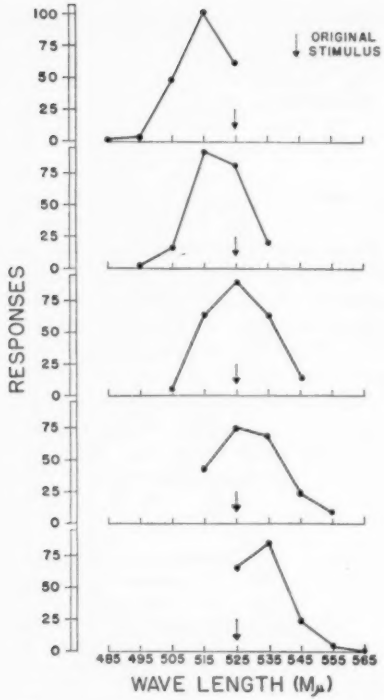

FIG. 1. Generalization gradients of the five experimental groups. (The gradient of Group 1 is at the top of the figure with that of Group 2 directly below it, etc. Note that the value of the original stimulus is the same for all groups.)

Procedure.—The *S*s were divided unsystematically into five groups of 10 *S*s each. Each *S* received the same instructions. They were as follows:

> This is an experiment in color perception. At the beginning of the experiment a specific color will be presented through the small hole in front of you. Try to keep this color in mind because you will be asked to identify it later. After 1 minute this color will be turned off and you will place your finger and press down on the telegraph key in front of you. I will give the signal "ready" and a few seconds later a color will again be presented. You must decide whether this is the original color shown you at the start of the experiment. If it is, *lift* your hand as rapidly as you can from the key. If it is not, keep pressing on the key.
>
> I will say the word "ready" whenever I am about to present a color and you should be pressing the key at that time. We are going to try some practice trials. Now we are going to run through a series exactly as we would do if this were the real experiment.

At this point the instructions were interrupted and a stimulus of 600 mμ was presented for 1 min. Then *S* was presented with 610 mμ, 590 mμ, 600 mμ, 620 mμ, and 580 mμ. Each test stimulus was presented for 5 sec. with 5 sec. intervening between presentations. If *S* appeared to have understood the instructions, they were continued as follows:

> Now we are going to begin the experiment. Remember, try to keep the original color in mind and respond as rapidly as you can, lifting your finger only when the original color appears. Do not be disturbed, however, if you should respond to other colors.

The *S*s in all five groups were presented with the same original stimulus, 525 mμ (a middle-green), for 60 sec. The groups differed only with regard to the series of stimuli employed in testing for generalization. In Group 1 the test stimuli were 485 mμ through 525 mμ, in 10-mμ steps. For Group 2 the range covered was 495 mμ–535 mμ, for Group 3, 505 mμ–545 mμ, for Group 4, 515 mμ–555 mμ, and for Group 5, 525 mμ–565 mμ. For each *S* the five test stimuli were randomized within a series and 12 different series were presented. The number of responses made to the different test stimuli constituted a generalization gradient.

RESULTS AND DISCUSSION

In Fig. 1 are presented the generalization gradients of the five groups of *S*s. It should be remembered that all groups were exposed to the same standard stimulus. In spite of this, the five gradients differ strikingly in a manner consistent with the findings of Philip (1952). The tendency to respond to stimuli closer to the center of the test series is so strong that the peak of the generalization gradient tends to be displaced from

the value of the original stimulus. Thus, a change of, for example, 10 mμ from the original stimulus does not produce a fixed generalization decrement, but may result in no change in response strength or even in an increment depending on the location within the range of test stimuli. It should be noted that only with Group 3, where the test stimuli were symmetrically distributed around 525 mμ, did the peak of the gradient fall clearly at that value!

The difference in gradient shape was shown to be statistically reliable in the following manner: A simple analysis of variance was performed to test for differences in the mean number of responses given to 515 mμ in Groups 1, 2, 3, and 4. The result was an F of 9.06 ($df = 3/36$, $P < .01$). A parallel analysis was performed for mean responses to 535 mμ in Groups 2, 3, 4, and 5. The result was an F of 8.41 ($df = 3/36$, $P < .01$).

The purpose of this study was to determine whether the effect of "central tendency" on the generalization gradient is peculiar to the absolute judgment situation employed by Philip. It is safe to conclude that it is not. Indeed, the distortion of the gradient proved even greater in the present experimental situation than under the conditions of Philip's experiment.

These findings may be interpreted with reference to Helson's (1947) theory of adaptation level. It may be argued that the series of test stimuli provides a frame of reference against which the memory trace of the original stimulus is judged. When test stimuli are presented which fall asymmetrically around the original stimulus, a change in the frame of reference may be assumed to result, culminating in a heightened tendency to respond to stimuli nearer to the center of the test series, thereby distorting the resulting generalization gradient.

The significance of the "central tendency effect" for studies of generalization employing conditioning techniques remains an open question. We would guess, however, that the background of test stimuli would exert far less influence in that situation than in the present one. In this study, *S*'s experience with the original stimulus was limited to one 60-sec. exposure. The resulting limited familiarity with the stimulus makes extensive retroactive interference by the test series which follows more likely. In a conditioning situation, however, *S* has the opportunity through repeated exposure to become much more familiar with the value of the CS. Greater familiarity with the stimulus should reduce the effect of the test situation which is later employed. Some direct evidence on this issue has been reported. Guttman (1956) discussed some pilot work with pigeons and an operant conditioning technique in which the asymmetry of the distribution of test stimuli appeared to have no effect on the resulting generalization gradient. More recent work from the Kent State laboratory has tended to corroborate his finding. Of course the species-related and procedure-related differences between the pigeon and human studies are so great as to make these findings merely suggestive. Research is needed to assess the role of familiarity with the CS in determining the influence of the "central tendency effect" on generalization.

Summary

Five groups of 10 *S*s each viewed a monochromatic light of 525 mμ (a middle-green) for 60 sec., and then were presented 12 different random series of wave lengths under instructions to respond only to the original color. The number of responses made to the different test stimuli constituted a gradient of generalization. Group 1 was tested with the series 485–525 mμ, in 10-mμ steps; Group 2, 495–535 mμ; Group 3, 505–545 mμ; Group 4, 515–555 mμ; and Group 5, 525–565 mμ. Only Group 3, with a central value of 525 mμ, produced a generalization gradient with a definite peak at 525 mμ; in all other cases the peak of responding was displaced toward the center of the series of test stimuli. The extent

of this displacement varied directly with the degree of asymmetry of the test series around the value of the original stimulus. These results support the assumption that the generalization test series serves as a frame of reference against which the memory trace of the original stimulus is judged.

References

DVORINE, I. *Dvorine color perception testing charts.* Baltimore, Md.: Waverly, 1944.

GUTTMAN, N. The pigeon and the spectrum and other perplexities. *Psychol. Rep.*, 1956, **2**, 449–460.

HELSON, H. Adaptation level as a frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60**, 1–29.

HOLLINGWORTH, H. L. The inaccuracy of movement. *Arch. Psychol., N. Y.*, 1909, **2** (Whole No. 13).

KALISH, H. I. The relationship between discriminability and generalization: A re-evaluation. *J. exp. Psychol.*, 1958, **55**, 637–644.

PHILIP, B. R. Effect of length of series upon generalization and central tendency in the discrimination of a series of stimuli. *Canad. J. Psychol.*, 1952, **6**, 173–178.

(Received June 28, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 1, 81–86

EFFECT OF REWARD MAGNITUDE, PERCENTAGE OF REINFORCEMENT, AND TRAINING METHOD ON ACQUISITION AND REVERSAL IN A T MAZE[1]

WINFRED F. HILL
Northwestern University
JOHN W. COTTON
University of California, Santa Barbara
AND KEITH N. CLAYTON
Vanderbilt University

The effect of reward magnitude on extinction is an unresolved question. Zeaman (1949) and Metzger, Cotton, and Lewis (1957) have found faster running early in extinction for animals receiving the larger acquisition reward, but convergence of the curves as extinction progressed. This is consistent with the assumption (Hull, 1952; Spence, 1956) that the incentive motivation factor (K) adjusts rapidly but not immediately to a change in reward magnitude. On the other hand, Hulse (1958) and Armus (1959) have found faster running over a number of extinction trials for animals receiving the smaller reward. This might be predicted from the depression effect found in some studies of magnitude change during acquisition (Spence, 1956), since no reward is a greater reduction from large reward than from small, but the generality of this effect is open to question. All of the above findings were with 100% reinforcement in acquisition. Hulse's study further complicated the picture by showing an interaction between magnitude and percentage of reward, with large reward producing greater resistance to extinction than small if rewards were intermittent during acquisition.

The present study translates the question of the effect of reward magnitude on extinction into a special case of extinction—discrimination reversal. In particular, it tests whether the interaction between magnitude and percentage of reward found by Hulse in straight-alley extinction will also appear in reversal of a T maze habit.

While making this test, it is possible in the same experiment to study the effect of reward magnitude on acquisition. A number of prior experiments suggest that larger rewards lead to faster learning of a discrimination if the noncorrection method is used but not if the correction method is used, either in a single-choice situation or in a multiple-unit maze. McKelvey (1956) investigated this apparent interaction of magnitude and method directly and found no effect of magnitude on correct choices with either method. However, he manipulated magnitude by varying the time that rats were permitted to eat rather than the amount of food available. Since rats learn to speed up their eating over a period of deprivation, it is possible that his "large reward" and "small reward" groups may have

[1] This study was carried out at Northwestern University and was supported by Grant G8706 from the National Science Foundation.

been eating nearly equal amounts at the end of acquisition. This possibility is strengthened by the fact that an initial difference in speed between the two groups disappeared by the end of acquisition. The acquisition phase of the present experiment incorporates the same basic design as McKelvey's, but with reward magnitude defined by quantity rather than by eating time.

METHOD

Subjects.—The *S*s were 96 female albino rats of the Sprague-Dawley strain purchased from Holtzman Rat Company, Madison, Wisconsin. Eight additional *S*s were discarded because of failure to run, refusal to eat the reward pellets, or *E*'s error. The *S*s were between 72 and 86 days old at the beginning of training.

Apparatus.—The *S*s were trained in an enclosed, single-unit T maze. This consisted of a start box 10 in. long and 6 in. wide, a stem 48 in. long and 4 in. wide, arms 7 in. long and 4 in. wide, and goal boxes 13 in. long and 6 in. wide. An alcove for food opened off each goal box in such a way that the reward was not visible from the entrance to the goal box. The entire maze was 8¼ in. high, with sides and floor of plywood and lids of Plexiglas. The inside was painted flat black. A guillotine door separated the start box from the stem, and sliding doors separated the stem from the arms. A Standard Electric timer was started by *S*'s weight on a treadle just beyond the start box and stopped by *S*'s weight on either of two treadles just beyond the guillotine doors.

Design.—All *S*s received 48 acquisition trials followed by 24 reversal trials. The design was a factorial combination of four dichotomous variables: (*a*) percentage reinforcement of the correct side during acquisition (100% vs. 50% random), (*b*) magnitude of reward during acquisition (four pellets vs. one), (*c*) acquisition training method (correction vs. forced-trial noncorrection), and (*d*) magnitude of reward during reversal. Thus there were 8 treatment cells of 12 *S*s in acquisition and 16 cells of 6 *S*s in reversal. Half of the *S*s in each cell were initially trained to the right side, half to the left.

It was desirable to hold both number of acquisition trials and proportional number of acquisition reinforcements constant for both magnitudes and both probabilities of reward. For noncorrection *S*s, this required the use of forced trials. Hence, noncorrection *S*s received a combination of 16 free and 32 forced trials in acquisition, so arranged that half of their acquisition responses were to each side. A given ordinal-numbered trial was either free or forced for all noncorrection *S*s, but each *S* in a given cell had a different sequence of forced right and left turns. All trials by the correction method were free. Reversal training for all *S*s was by the noncorrection method with all trials free and with 100% reinforcement.

The above procedure had the effect that a correction *S* received twice as many reinforcements in acquisition as a noncorrection *S* in the corresponding group. However, since method was of interest in connection with acquisition rather than with reversal, this difference was not crucial. Experiments on the role of incorrect responses in discrimination learning suggest that this procedure should make correction and noncorrection groups more nearly comparable than if number of reinforcements were held constant and number of incorrect responses allowed to vary.

Procedure.—Seven days before the beginning of training, *S*s were placed on a feeding schedule of 10 gm. of ground Purina chow a day, which continued throughout the experiment. On 6 of these 7 preliminary days, each *S* received 3 min. of handling and was allowed to eat four of the reward pellets. Water was always present in the living cages and the carrying cages throughout the experiment.

In both acquisition and reversal, *S*s were given six trials a day. Thus there were 8 days of acquisition and 4 of reversal, successive except for 1 nontraining day between Days 6 and 7 of acquisition. The reward pellets were 45-mg. Noyes pellets. At the start of each trial *S* was placed in the start box. When *S* was oriented toward the door, *E* opened it, permitting *S* to enter the stem. In the noncorrection procedure, after *S* entered either arm the sliding door was closed and *S* was confined for approximately 15 sec. (If it was a forced trial, the door to the other arm was already closed.) In the correction procedure, if *S* entered the incorrect arm the door remained open and *S* was permitted to retrace, but when *S* entered the correct arm the door was closed and *S* was confined for 15 sec. In both conditions *S* remained in a carrying cage for about 30 sec. between trials. The *S*s received their daily ration about ½ hr. after training.

Results

Acquisition.—The number of correct choices was tabulated for each *S* in acquisition. Since only the free trials were relevant for the noncorrection *S*s, only the corresponding ordinal-numbered trials were counted for the correction *S*s. Thus each acquisition score represented the number of correct turns out of a possible 16. Three cases of failure to run (out of 1536 trials) were counted as errors.

The effects of the three acquisition variables are shown in the three pairs of curves in Fig. 1, each curve based on 48 rats. A triple-classification analysis of variance, with $df = 1/88$ for each F ratio, indicated that all three main effects were significant. Acquisition was faster for four pellets than for one ($F = 6.71$, $P = .05$), for 100% reinforcement than for 50% ($F = 16.86$, $P = .001$), and for forced noncorrection method than for correction ($F = 7.38$, $P = .01$). None of the interactions was significant, all four Fs being less than unity. The Hartley test did not indicate any heterogeneity of variance ($F_{max.}$ of 3.64). Although there was a marked negative skew in the distribution of scores, it is unlikely (Boneau, 1960) that this had any substantial effect on the Fs.

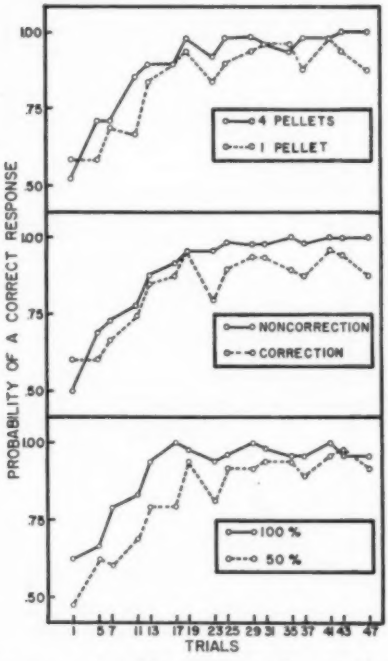

FIG. 1. Acquisition for all *S*s, classified according to the three acquisition variables.

TABLE 1

NUMBER OF CORRECT CHOICES IN REVERSAL

% Reinf.	Acquisition Reward	Method	Reversal Reward	Mean	*SD*
100	1	C	1	14.83	6.58
100	1	C	4	19.67	1.36
100	1	N	1	10.17	1.94
100	1	N	4	14.00	1.90
100	4	C	1	17.50	2.35
100	4	C	4	17.83	1.75
100	4	N	1	8.50	1.87
100	4	N	4	14.33	3.26
50	1	C	1	17.17	4.02
50	1	C	4	18.17	2.40
50	1	N	1	12.17	3.97
50	1	N	4	18.50	2.07
50	4	C	1	7.67	4.08
50	4	C	4	10.00	5.29
50	4	N	1	7.83	5.34
50	4	N	4	10.17	5.23

Note.—$N = 6$ each group.

Reversal.—The number of correct choices on the 24 reversal trials was tabulated for each *S*. Table 1 presents the means and *SD*s of these scores for the 16 groups. These distributions were more nearly normal than those for acquisition, and the apparent heterogeneity of variance was not significant by Bartlett's test ($B = 19.76$, $df = 15$).

Table 2 summarizes the quadruple-classification analysis of variance of these data. Although all four main effects were significant when tested

TABLE 2

ANALYSIS OF VARIANCE OF NUMBER OF CORRECT CHOICES IN REVERSAL

Source	*df*	*MS*	*F*
Percentage reinforcement (%)	1	84.376	6.20*
Training method (M)	1	273.376	20.07***
Acquisition reward (*KA*)	1	360.376	26.46***
Reversal reward (*KR*)	1	266.667	19.58***
% × M	1	126.040	9.70**
% × *KA*	1	330.040	24.23***
% × *KR*	1	2.667	—
M × *KA*	1	3.374	—
M × *KR*	1	37.499	2.75
KA × *KR*	1	10.667	—
% × M × *KA*	1	18.377	1.35
% × M × *KR*	1	.167	—
% × *KA* × *KR*	1	0	—
M × *KA* × *KR*	1	.667	—
% × M × *KA* × *KR*	1	53.998	3.96
Within cells[a]	80	13.620	

[a] Used as error term for all *F* tests.
* $P < .05$.
** $P < .01$.
*** $P < .001$.

against the within-cells variance, the three that involved acquisition variables were overshadowed by interactions. Figure 2 shows the course of reversal learning classified in three different ways so as to clarify these interactions. The effect of acquisition variables on reversal may be summarized by saying that 100% correction *Ss* reversed fastest and 50% large-reward *Ss* reversed most slowly. Reversal, like acquisition, was faster for the larger current reward.

The number of errors prior to the first correct response in reversal was also tabulated and subjected to an analysis of variance on the three acquisition variables. This measure gave substantially the same results as the total correct reversal responses. The main effects of percentage and method fell short of significance, but the effect of acquisition magnitude remained significant at the .001 level. The interactions of percentage with reward size and percentage with method manifested the same patterns as before and were both significant at the .01 level.

Speeds.—Time scores were converted to reciprocals and analyses of variance computed. In a triple-classification analysis for the last four trials of acquisition, with 1/88 *df* for each *F*, the correction method gave faster running at the .001 level ($F = 131.43$) and larger reward gave faster running for correction-method *Ss* only (*F* for magnitude-method interaction = 4.70, $P = .05$). No other *F* was significant. In a quadruple-classification analysis for all 24 reversal trials, with 1/80 *df* for each *F*, *Ss* receiving the larger current reward ran faster ($F = 10.16$, $P = .01$) and those trained with the correction method ran faster ($F = 8.90$, $P = .01$). No other *F* was significant.

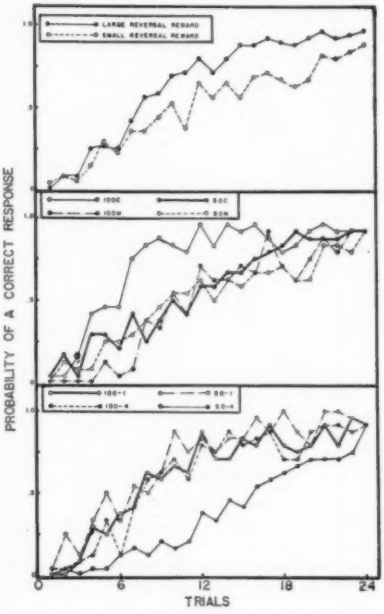

FIG. 2. Reversal learning for all *Ss*, classified in three different ways according to reversal variable (top) and acquisition variables (middle and bottom).

Discussion

Acquisition.—The superiority of the forced-trial, noncorrection method, in spite of the smaller number of reinforcements that it provided, argues for the importance in learning of making nonrewarded incorrect responses. Not only did the correction *Ss* have less experience with the incorrect side, they also were able to obtain reward, although delayed, even when they made the wrong choice. The fact that the noncorrection *Ss* were forced half the time to the wrong side makes it difficult to compare these results with those of other studies comparing the correction and noncorrection methods. However, in emphasizing the importance of incorrect responses, this study agrees with others that have found better learning of a discrimination with a combination of rewarded and nonrewarded trials than with the same number of trials all rewarded.

No support was found for the hypothesis that greater reward leads to faster spatial discrimination learning with the noncorrection but not with the correction method. Larger reward gave faster learning throughout, and there was no interaction between magnitude and method. The possibility that the present finding is due to the use of forcing in the noncorrection method cannot be ruled out. However, since this experiment was begun, another finding disconfirming the hypothesis has been reported by Lawson, Cross, and Tambe (1959). Thus it appears that some other factor must be found to explain the discrepant prior findings on the relation of discrimination learning to reward magnitude.

The more rapid acquisition with 100% than with 50% reinforcement is comparable to the common, though by no means universal, finding with acquisition of a simple running response.

Reversal.—The finding that a large reward 50% of the time produced greater resistance to reversal than any of the other three combinations of magnitude and probability is similar to Hulse's (1958) finding on resistance to extinction in a straight alley. However, the lack of any difference among the other three groups differs from Hulse's findings. The effect found here is consistent both with a partial reinforcement extinction effect, but only for the large reward condition, and with a greater persistence of more strongly rewarded responses, but only for the 50% condition, and hence is not completely consistent with either of these principles.

Training method had no appreciable effect on reversal when total number of reinforcements was held constant by comparing 50% correction with 100% noncorrection *Ss*. The interaction that was found between method and percentage may reflect in part a tendency to prefer the less often experienced side (Denny, 1957). Since the 100% correction *Ss* made the largest number of correct turns in acquisition (the noncorrection *Ss* being forced equally often to the two sides), this tendency should be greatest for them and should lead to more rapid extinction of the old response and acquisition of the new response, as was found. The partial reinforcement effect might then be invoked to explain why the 50% correction *Ss* were so markedly slower in reversing than the 100% correction *Ss*. According to this combination of factors, however, the 50% noncorrection group should have reversed most slowly, which was not the case. Thus this interaction also remains at least partly unexplained.

The faster reversal with larger reversal reward agrees with the finding of faster acquisition with larger reward. Since the reversal training method of free, noncorrection trials was different from either of the acquisition methods, the superiority of the larger reward in this condition increases the generality of the finding that larger reward leads to faster discrimination learning. There was no evidence of generalization decrement resulting from the change in reward magnitude between acquisition and extinction. The lack of interaction between reversal reward magnitude and any of the acquisition variables argues for the independence of acquisition and

reversal variables in their effect on reversal.

Speeds.—Since the recorded running speeds include time in the stem plus time in the choice area, the speed data are difficult to interpret. The most striking finding is the superiority of the correction method. McKelvey (1956) also found greater speed and less accuracy with the correction method, and in his study this might be attributed to the greater consistency of reward with correction method. This explanation seems inadequate in the present study, however, since 100% reinforcement did not give significantly faster running than 50%. It is also striking that the superiority of the correction group continues through reversal, even though all *S*s were on noncorrection method during reversal. Since the noncorrection *S*s received primarily forced trials in acquisition, it seems likely that forcing to the incorrect side, rather than noncorrection as such, was the crucial factor. However, why the difference should remain after all animals were changed to free trials is problematic.

In general the effect of magnitude is greater on choices than on speeds. In particular, noncorrection *S*s made more correct choices in acquisition, but did not run faster, for large than for small reward. This argues against Pubols' (1961) suggestion that the effect of magnitude on choices is mediated by the effect on speeds.

Summary

Acquisition and reversal in a T maze were studied for 96 rats as a function of four variables combined factorially: (*a*) 100% vs. 50% random reinforcement in acquisition, (*b*) one vs. four reward pellets in acquisition, (*c*) correction vs. forced-trial noncorrection method in acquisition, and (*d*) one vs. four reward pellets in reversal. All *S*s received 100% reinforcement and free-trial noncorrection method in reversal.

Acquisition of the correct response was faster for large reward, 100% reinforcement, and forced noncorrection method, with no interactions. Reversal was faster for large reversal reward, faster after 100% reinforcement with correction method than any other combination of percentage with method, and slower after 50% reinforcement with large acquisition reward than any other combination of percentage with acquisition magnitude. Acquisition by the correction method gave faster running during both acquisition and reversal. Running was faster with larger current reward both during acquisition by correction method and during reversal.

References

Armus, H. L. Effect of magnitude of reinforcement on acquisition and extinction of a running response. *J. exp. Psychol.*, 1959, **58**, 61–63.

Boneau, C. A. The effects of violations of assumptions underlying the *t* test. *Psychol. Bull.*, 1960, **57**, 49–64.

Denny, M. R. Learning through stimulus satiation. *J. exp. Psychol.*, 1957, **54**, 62–64.

Hull, C. L. *A behavior system.* New Haven: Yale Univer. Press, 1952.

Hulse, S. H., Jr. Amount and percentage of reinforcement and duration of goal confinement in conditioning and extinction. *J. exp. Psychol.*, 1958, **56**, 48–57.

Lawson, R., Cross, H. A., & Tambe, J. T. Effects of large and small rewards on maze performance after different prior experiences with reward amounts. *J. comp. physiol. Psychol.*, 1959, **52**, 717–720.

McKelvey, R. K. The relationship between training methods and reward variables in brightness discrimination learning. *J. comp. physiol. Psychol.*, 1956, **49**, 485–491.

Metzger, R., Cotton, J. W., & Lewis, D. J. Effect of reinforcement magnitude and of order of presentation of different magnitudes on runway behavior. *J. comp. physiol. Psychol.*, 1957, **50**, 184–188.

Pubols, B. H. The acquisition and reversal of a position habit as a function of incentive magnitude. *J. comp. physiol. Psychol.*, 1961, **54**, 94–97.

Spence, K. W. *Behavior theory and conditioning.* New Haven: Yale Univer. Press, 1956.

Zeaman, D. Response latency as a function of the amount of reinforcement. *J. exp. Psychol.*, 1949, **39**, 466–483.

(Received June 30, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 1, 87–93

PAIRED-ASSOCIATE LEARNING UNDER SIMULTANEOUS REPETITION AND NONREPETITION CONDITIONS[1]

WILLIAM F. BATTIG

University of Virginia

Several recent investigators (e.g., Clark, Lansford, & Dallenbach, 1960; Estes, 1960; Estes, Hopkins, & Crothers, 1960; Rock, 1957; Rock & Heimer, 1959) have obtained evidence leading them to conclude that associative learning is an all-or-none rather than a gradual incremental process, and that repetition serves only to provide additional opportunities for such all-or-none associations to be learned. However, none of these studies have been sufficiently free of defects in experimental design and procedure to preclude alternative interpretations, discussed at length elsewhere (Postman, 1962; Underwood, 1961; Underwood, Rehula, & Keppel, 1962), so that the theoretical issue of all-or-none vs. gradual association formation remains a matter of considerable dispute.

The present paper reports a series of experiments designed to eliminate or minimize the influence of several inadequately controlled variables in the procedure originally introduced by Rock (1957), wherein performance of a repetition group receiving the same list of paired-associate items on each trial did not differ from a modified nonrepetition group, for whom all pairs not correctly responded to on any trial were removed from the list and replaced by new pairs on the following trial. Since groups learning under the repetition and nonrepetition procedures may also differ in such factors as instructions and approach to the task (Brackett, 1961), or amount of interference produced by other pairs in the list (Brown, 1961; Clark et al., 1960), a within-*Ss* design was employed so that each *S* learned simultaneously under both conditions and therefore served as his own control. To test more adequately the role of repetition in association formation as distinguished from other processes involved in paired-associate learning (e.g., Underwood & Schulz, 1960), an attempt was made to develop materials which provided a relatively pure and substantial case of association formation. These materials were also highly homogeneous and carefully calibrated so as to be equivalent in difficulty, in order to minimize the effects of selective elimination of the more difficult incorrect pairs on each trial, which has been shown to result in a significantly easier list under the nonrepetition condition (Underwood et al., 1962; Williams, 1961).

Experiments I and II

Method

The method and procedure for both Exp. I and II closely followed Rock (1957), except that the list for each *S* consisted of two equal-sized subsets of pairs representing the repetition (R) and nonrepetition (NR) conditions. All pairs of the R subset were presented on every trial throughout the experiment, whereas only those pairs responded to correctly on the preceding trial were retained in the NR subset, all incorrect pairs being removed from the list and replaced on the next trial

[1] Experiments I and II were reported at the April 1959 meetings of the Eastern Psychological Association, Atlantic City, N. J.; Exp. III and IV were supported by a contract with the United States Office of Education, Department of Health, Education, and Welfare. The author is indebted to Douglas Nelson for assistance with Exp. IV.

by new pairs. The 15 *S*s in each experiment were given typical paired-associate instructions with an added statement that new pairs might be introduced during the experiment, although the procedure effectively prevented *S* from distinguishing between the R and NR subsets, or from detecting the basis for introducing new NR pairs.

All learning materials were typed on 3×5 in. cards and presented manually through a card-exposure device. Each trial began with the successive presentation of all pairs in the list for 3 sec. each with a 5-sec. interpair interval, followed by all stimulus terms alone in a different order at a 5-sec. rate while *S* attempted to recall the associated response, with a 30-sec. intertrial interval. To further guard against differentiation of the R and NR subsets, both were distributed evenly throughout the list on each trial.

Experiments I and II differed only in the kind of learning material employed. In Exp. I, 84 pairs of nonsense syllables of 47–53% association value (Glaze, 1928) were selected so as to be maximally homogeneous in judged ease of learning according to a previous scaling procedure (Battig, 1959), and divided into 14 subsets of 6 pairs, matched as to mean and *SD* of these scale values. The initial list of 12 pairs for each *S* consisted of two of these equivalent 6-pair subsets representing the R and NR conditions, replacement pairs for the latter being selected from another equivalent subset until exhausted, then from another subset, and so on until the final trial. Each of the 14 subsets was used equally often as the R and initial NR subset and at each point in the sequence of replacement pairs for the various *S*s. Learning was carried out to a criterion of 1 errorless trial or a maximum of 10 trials.

Pairs of common 4-letter words as stimuli and 2-digit number responses were used in Exp. II, wherein each *S* learned to a criterion of one errorless trial an 18-pair list consisting of two 9-pair subsets of R and NR pairs. A total of 84 such word-number pairs were constructed so as to avoid obvious intrapair associations or similarities, each being used equally often in the R and initial NR subset and at various points in the sequence of NR replacements.

Each *S* in both Exp. I and II was given a recognition test immediately after the last trial, consisting of the individual presentation of 10 incorrect NR pairs which had been replaced, intermixed randomly with 10 additional pairs from the pool of 84 which *S* had never seen, with *S* indicating in each case whether or not he thought he had seen the pair previously during the experiment.

Results

Means and *SD*s for the repetition (R) and nonrepetition (NR) conditions are presented separately in Table 1 for total errors over the six trials on which data were available from all 15 *S*s of Exp. I, and for total errors and trials to the criterion of one errorless trial in Exp. II. Revealed herein is a marked and statistically significant difference in favor of Cond. R in Exp. I ($t = 6.44$, $df = 14$), whereas the slight superiority of Cond. NR for both measures in Exp. II fell far short of significance (both t's < 1). However, the recognition test showed significantly above-chance recognition of incorrect NR pairs which was actually superior in Exp. II (79.7% correct, $t = 7.76$) to Exp. I (70% correct, $t = 7.75$), demonstrating that something had been learned about the replaced incorrect pairs even in Exp. II where their removal did not retard NR learning.

The results of Exp. I, wherein every 1 of the 15 *S*s made more NR than R errors, may be due to facilitation by repetition of the substantial response learning required for the nonsense-syllable materials, and therefore are inadequate to disprove an all-or-none theory of association formation as distinguished from paired-associate learning (e.g., Underwood et al., 1962). Although such response learning was minimized for the word-number pairs of Exp. II, so also was the amount of association formation significantly reduced, resulting in such rapid learning (63% correct responses on Trial 2) that the performance measures become quite insensitive to any effects of repetition. Since these materials also differed considerably in difficulty, thereby providing an important advantage for Cond. NR, the results of Exp. II cannot be interpreted as unequivocally supporting an all-or-none theory of association formation.

Experiment III

To overcome the major deficiencies of Exp. I and II, an attempt was made to develop for use in Exp. III materials which required a substantial amount of association formation while relatively free from other nonassociative processes typically involved in paired-associate learning. This was accomplished by (*a*) constructing pairs with maximally different and unrelated stimulus and response members; (*b*) using stimuli of low meaningfulness and familiarity which were highly discriminable from each other, since previous evidence indicates that discriminability represents a major factor in learning of the stimulus (Battig, Williams, & Williams, 1962), whereas meaningfulness and familiarity are of lesser importance (Underwood & Schulz, 1960); (*c*) using highly meaningful and familiar responses to minimize response learning.

Method

Each of 72 nonsense shapes of high discriminability and low association value was paired with a different two-digit number between 12 and 98 (excluding double numbers and numbers ending in zero), forming 72 pairs with minimal shape-number similarity according to the judgments of 10 preliminary *Ss*. Previous measures had been obtained of (*a*) association value and discriminability of each shape from 122 *Ss*; (*b*) association value of each number from 28 *Ss*; (*c*) rated "ease of learning" of each shape-number pair by 66 *Ss*; (*d*) actual learning difficulty in a separate study reported elsewhere, which describes the materials in more detail (Battig & Brackett, 1961). These 72 pairs were divided into 12 subsets of 6 pairs each which were approximately equivalent with respect to each of these four indices, each being used with approximately equal frequency as the R and initial NR subset and throughout the sequence of NR replacement pairs.

All other aspects of method and procedure were identical to Exp. I with the following minor exceptions: (*a*) all materials were photographed and presented for learning and attempted recall by means of an automatic slide projector; (*b*) a 5-sec. rate (4-sec. exposure, 1-sec. interslide interval) was used for presentation of pairs for learning as well as for stimuli alone during attempted recall, with a 30-sec. interval between trials and between presentation and recall series within each trial; (*c*) learning was continued to a criterion of two successive errorless trials (although none of the 15 *Ss* ever made an error on the trial following the first errorless trial); (*d*) following the postexperimental recognition test for incorrect NR pairs, each *S* was shown again each of the 12 pairs he had learned and asked to describe verbally how he had learned them.

TABLE 1

Performance Measures for Repetition and Nonrepetition Conditions of Exp. I–III

Exp.	Measure	Repetition		Nonrepetition	
		Mean	*SD*	Mean	*SD*
I	Errors	22.13	5.40	28.93	5.25
II	Errors	13.27	6.95	12.40	5.87
	Trials	4.47	1.41	4.20	1.17
III	Errors	16.00	4.87	20.07	7.04
	Trials	5.80	1.17	6.67	1.85

Note.—Errors and trials are to a criterion of one errorless trial, except for Exp. I which is based on Trials 1–6 for all *Ss*.

Results

As shown in Table 1, performance was superior on the R subset according to both error and trial measures, although significantly so only for errors ($t = 2.83$, $P < .02$) but not for trials ($t = 1.85$, $.05 < P < .10$). Further error analysis revealed the difference to be primarily due to significantly more response omissions under Cond. NR ($t = 3.06$, $P < .01$), which accounted for 29% of the NR errors as compared with 18% for Cond. R. The two conditions did not differ significantly either in partially correct responses ($t = 1.25$), intralist intrusions ($t = 1.34$), or extralist intrusions ($t = 1.05$). Recognition test results were similar to Exp. I and II, yielding 71.7% correct identifications

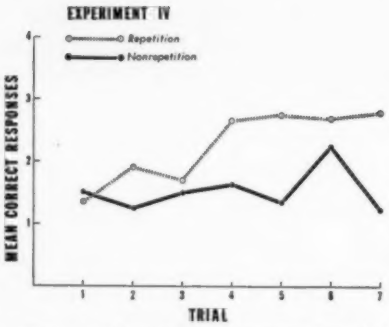

FIG. 1. Mean correct responses per trial for the modified repetition and nonrepetition conditions of Exp. IV on Trials 1–7.

($t=9.12$). Further analysis revealed recognition errors to consist primarily of reports of incorrect NR pairs as not seen previously (85%) rather than of the unseen pairs as previously seen (15%).

Based on postexperimental questioning, the learned pairs were categorized on the basis of *S*'s ability to verbalize a mediated basis for associating the shape and number together, yielding insignificantly more mediated associations for NR (24%) than R pairs (20%). No basis for association could be verbalized for 67% and 63% of the R and NR pairs, respectively, the remaining 13% in each case representing a "doubtful" category. These results indicate that the present materials had been reasonably successful in providing a relatively pure and substantial case of association formation, and were sufficiently homogeneous to minimize if not eliminate the problem of differences in pair difficulty in favor of Cond. NR.

EXPERIMENT IV

Although the differences in favor of R over NR performance in Exp. III were significant statistically, they were not impressively large in magnitude, due at least in part to the relatively small number of pairs learned by each *S*, and the predominance of already learned pairs on all but the first few trials as a consequence of the rapid learning of the shape-number pairs. In order to overcome these deficiencies and provide a more sensitive test of the role of repetition in this task, the procedure was modified in Exp. IV so that all pairs learned on each trial were eliminated from the list and replaced by new pairs.

Method

As in Exp. III, each *S* was presented with a list of 12 shape-number pairs on each trial, divided into two equivalent subsets of 6 pairs representing repetition (R) and nonrepetition (NR) conditions. However, all NR pairs were presented only for a single trial, being replaced after each trial by another of the equivalent subsets of 6 pairs which *S* had not seen previously. Pairs of the R subset were removed from the list and replaced by new pairs only if they had been responded to correctly on the preceding trial, so that this subset on any given trial typically included a combination of new pairs with previously presented incorrect pairs. Trials were continued in this manner until the pool of 72 pairs was exhausted, which required seven–nine trials depending on the number of correct responses to R pairs for the various *S*s.

The only other differences in method and procedure from Exp. III consisted of (*a*) appropriate modifications in instructions to cover the changed conditions; (*b*) elimination of the postexperimental recognition test and attempted verbalization of correct associations; (*c*) use of 20 paid *S*s representing a wider range of previous experience and sophistication about paired-associate learning experiments than in Exp. I–III, each of which had used 15 volunteers from undergraduate psychology courses.

Results

The mean number of correct responses per trial for pairs of the R and NR subsets are presented in Fig. 1, which indicates a consistent and increasing superiority of Cond. R to NR over the seven trials on which data were available from all *S*s.

Total correct responses for Cond. R (Mean = 17.85, SD = 4.24) were significantly greater (t = 7.06, df = 19) than for Cond. NR (Mean = 11.65, SD = 4.96). Except for 2 Ss who performed identically under both conditions, all Ss made more correct responses on the R subset. The slopes of the R and NR curves of Fig. 1 were found to differ significantly by trend analysis of variance (F = 10.35, df = 1/38), due to a highly significant improvement over trials for Cond. R (F = 29.92, df = 1/19) whereas the slight increase in correct responses under Cond. NR fell far short of significance (F < 1). In agreement with Exp. III, error analysis revealed significantly more omissions (t = 2.86, P < .01) for Cond. NR (Mean = 11.10) than Cond. R (Mean = 7.50), but the excess of partially correct responses and intrusions in Cond. NR (Mean = 24.35) over Cond. R (Mean = 21.75) was not significant (t = 1.58, P > .10).

Although the greater increase in correct responses over trials under Cond. R undoubtedly reflects the increased proportion of frequently presented incorrect pairs on later trials, a more direct and convincing demonstration of the superiority of performance on repeated pairs comes from a comparison of proportions of correct responses for R pairs varying in number of prior presentations. Table 2 presents the total number of pairs presented once, twice, etc. within the R subset summed over all Ss, along with the total number and proportion of these presentations which yielded correct responses, and the mean proportion of correct responses averaged over Ss[2]. Revealed herein is a consistent increase with number of prior presentations in both total and mean proportion of correct responses, the rate of increase being slightly less for total proportion due to the relatively greater contribution of slower learners to this figure as number of presentations increase. The proportions of first, second, and third presentations responded to correctly by each S were subjected to trend analysis of variance, which showed the increase with number of presentations to be highly significant (F = 24.56, df = 1/19).

TABLE 2

PROPORTIONS OF CORRECT RESPONSES FOR REPETITION PAIRS AS A FUNCTION OF NUMBER OF PRIOR PRESENTATIONS

Measure	Number of Prior Presentations				
	1	2	3	4	5+
Total number	420	265	135	66	56
Correct responses	118	114	61	32	28
Total proportion	.281	.430	.452	.485	.500
Mean proportion	.270	.424	.496	.581	.539

DISCUSSION

The present results demonstrate conclusively the facilitation of paired-associate learning by repetition of incorrect pairs under conditions where (*a*) other variables which may differ for R and NR conditions have been eliminated through the use of a within-Ss design; (*b*) careful precautions have been taken to minimize the effects of differences between pairs in difficulty; (*c*) nonassociative factors in paired-associate learning, such as response learning

[2] A somewhat more stable estimate of performance on once-presented pairs would probably be provided by the inclusion of pairs of the NR subset, all of which were presented only once. However, performance on NR pairs (p = .247) was somewhat below that for R pairs presented for the first time (p = .281), probably due to the differential distribution of these pairs over trials under the two conditions. The decision to base the present comparison solely on R pairs therefore represents, if anything, an overly conservative estimate of the magnitude of the increase in proportion of correct responses with number of previous presentations.

and stimulus discrimination, have been largely eliminated from the task. Only in Exp. II, which was clearly inadequate due to the extremely rapid learning of the word-number pairs and the lack of control for differences in pair difficulty, were the results not in direct conflict with an all-or-none theory, and even here the results of the recognition test showed that something short of a correct association had been learned about the eliminated incorrect NR pairs.

In comparison with previous studies using the Rock (1957) procedure, Exp. III and IV would appear to be clearly superior with respect to the elimination of uncontrolled variables and biases favoring either the R or NR condition. However, despite the extensive efforts to equate the shape-number pairs in average difficulty, these pairs probably were not equally difficult for the individual *S*. Furthermore, besides being more difficult, the increasing superiority of performance on more frequently presented R pairs in Exp. IV may have obtained in spite of the relatively large contribution of slow learners to measures based on these pairs. Since either or both of these sources of bias would clearly favor NR performance, the conclusiveness of the present results in demonstrating the positive role of repetition can only be enhanced thereby.

Only if the position is taken that some process other than association formation has been facilitated by repetition can the results of Exp. III and IV be regarded as insufficient to disprove an all-or-none theory of association formation. Any such argument, of course, reduces the issue to empirical meaninglessness unless the association-formation process can be given precise operational specification. While verbal definitions of association formation as distinguished from paired-associate learning suffer from considerable ambiguity, there seems to be reasonable agreement that this refers to the process of "connecting," "hooking-up," or "associating together" the stimulus and response members of each pair. Procedurally, efforts to provide the necessary pure case of association formation have concentrated on the elimination of response learning, reflecting the influence of Underwood's conception of paired-associate learning as a two-stage process consisting of response-learning and associative phases (Underwood et al., 1962; Underwood, Runquist, & Schulz, 1959; Underwood & Schulz, 1960). There can be little question but that the present results are directly applicable to this definition of association formation as paired-associate learning in the absence of response learning.

However, although the present stimulus materials were designed to be maximally discriminable, thereby eliminating another nonassociative process suggested to be involved in paired-associate learning (Battig et al., 1962; Gibson, 1940), they differed from the letters or words typically employed as stimuli in related studies in their low meaningfulness and familiarity. Inasmuch as stimulus familiarity has little or no effect on paired-associate learning while stimulus meaningfulness affects primarily the association-formation phase (Underwood & Schulz, 1960), the present shape-number materials would therefore appear to be maximally appropriate and sensitive as a test of the role of repetition in association formation, since they do not substantially reduce, nor shortcut through already existing mediated associations, the amount of association formation required of *S*.

Nevertheless, although the present materials have eliminated those nonassociative processes previously identified to be important in paired-associate learning, the possibility remains that other processes besides association formation are still involved in learning the shape-number pairs. However, unless or until such processes are identified and given adequate operational specification, it can be concluded that association formation as presently defined is not an all-or-none process, but instead builds up gradually in strength through repetition. Moreover, in view of the rapid learning observed even under the present conditions, which were carefully designed to maximize the required amount of

association formation, it would appear that this process may constitute such a small and insignificant part of the learning involved in the typical paired-associate task, that questions concerning its all-or-none or gradual nature are likely to be of little consequence for the general understanding of factors important in paired-associate learning.

Summary

The effect of repetition of previously incorrect pairs in paired-associate learning was evaluated in four separate experiments in which each *S* learned a single paired-associate list consisting of two equivalent subsets of pairs. In Exp. I–III, each of 15 *S*s learned under conditions where the repetition (R) subset consisted of the same pairs repeated on all trials, while pairs of the nonrepetition (NR) subset were retained in the list only if responded to correctly and were otherwise removed from the list and replaced by new pairs on the next trial. Lists of 12 nonsense-syllable pairs, 18 word-number pairs, and 12 pairs of nonsense shapes and numbers were used in Exp. I–III, respectively. Experiment IV ($N = 20$) used the same shape-number pairs as Exp. III, but under conditions where all pairs responded to correctly were immediately removed from the list, so that the NR subset consisted of a new set of pairs on each trial, while the R subset included both new pairs and previously presented incorrect pairs. Significant differences in favor of Cond. R were obtained in all cases except for Exp. II, which was attributed to the greater ease of learning and rather wide range of difficulty of the word-number pairs. Particularly in view of the significant positive effects of repetition in Exp. III and IV, using materials requiring a relatively pure and substantial case of association formation which were also carefully calibrated to minimize the effects of differences in pair difficulty, it was concluded that the present results are directly contradictory to an all-or-none theory of association formation in paired-associate learning.

References

Battig, W. F. Scaled difficulty of nonsense-syllable pairs consisting of syllables of equal association value. *Psychol. Rep.*, 1959, **5**, 126.

Battig, W. F., & Brackett, H. R. Comparison of anticipation and recall methods of paired-associate learning. *Psychol. Rep.*, 1961, **9**, 59–65.

Battig, W. F., Williams, J. M., & Williams, J. G. Transfer from verbal-discrimination to paired-associate learning. *J. exp. Psychol.*, 1962, **63**, 258–268.

Brackett, H. R. The effect of pretraining and knowledge of results on the acquisition of paired associates. Unpublished master's thesis, University of Virginia, 1961.

Brown, S. C. Interpair interference in paired-associate learning. Unpublished master's thesis, University of Virginia, 1961.

Clark, L. L., Lansford, T. G., & Dallenbach, K. M. Repetition and associative learning. *Amer. J. Psychol.*, 1960, **73**, 22–40.

Estes, W. K. Learning theory and the new "mental chemistry." *Psychol. Rev.*, 1960, **67**, 207–223.

Estes, W. K., Hopkins, B. L., & Crothers, E. J. All-or-none and conservation effects in the learning and retention of paired associates. *J. exp. Psychol.*, 1960, **60**, 329–339.

Gibson, E. J. A systematic application of the concepts of generalization and differentiation to verbal learning. *Psychol. Rev.*, 1940, **47**, 196–229.

Glaze, J. A. The association value of nonsense syllables. *J. genet. Psychol.*, 1928, **35**, 255–267.

Postman, L. Repetition and paired-associate learning. *Amer. J. Psychol.*, 1962, **75**, in press.

Rock, I. The role of repetition in associative learning. *Amer. J. Psychol.*, 1957, **70**, 186–193.

Rock, I., & Heimer, W. Further evidence of one-trial associative learning. *Amer. J. Psychol.*, 1959, **72**, 1–16.

Underwood, B. J. One-trial learning. Invited address at Midwestern Psychological Association, Chicago, May 1961.

Underwood, B. J., Rehula, R., & Keppel, G. Item selection in paired-associate learning. *Amer. J. Psychol.*, 1962, **75**, in press.

Underwood, B. J., Runquist, W. N., & Schulz, R. W. Response learning in paired-associate lists as a function of intra-list similarity. *J. exp. Psychol.*, 1959, **58**, 70–78.

Underwood, B. J., & Schulz, R. W. Meaningfulness and verbal learning. Chicago: Lippincott, 1960.

Williams, J. P. Supplementary report: A selection artifact in Rock's study of the role of repetition. *J. exp. Psychol.*, 1961, **62**, 627–628.

(Received July 3, 1961)

Journal of Experimental Psychology
1962, Vol. 64, No. 1, 94

SUPPLEMENTARY REPORT: EFFECTS OF STIMULUS ASSOCIATION VALUE AND EXPOSURE DURATION ON R-S LEARNING[1]

NED CASSEM AND DONALD H. KAUSLER

St. Louis University

Jantz and Underwood (1958) demonstrated a gross positive relationship between stimulus association value (AV) and the amount of R-S learning. However, they found the same reversal between 0% and 33% Glaze values that had been reported earlier by Postman, Adams, and Phillips (1955). As noted by Jantz and Underwood, the reversal most likely represented a sampling artifact in that only two syllables were included in their list at each AV level. The present study attempted to provide a more precise determination of the relationship between the AV of stimuli and R-S learning by means of a more adequate sampling of stimulus items.

In addition, the present study investigated the effects of exposure duration during both S-R learning and R-S recall on the amount of R-S recall. If, as proposed by Feldman and Underwood (1957), R-S learning is a variant of incidental learning, in which *S* learns associations for which he has no learning set during a period in which he is set to learn other associations, then the amount of R-S or incidental learning should increase with increased exposure duration (Kausler & Trapp, 1961; Rosenberg, 1959).

Method.—Eighty seminarians were randomly assigned to the 16 groups of a 2×2×4 factorial design, representing exposure duration during S-R learning (T_L), exposure duration during R-S recall (T_R), and AV. Each *S* had six trials on an eight-item list with nonsense syllables as stimuli and the same words employed by Jantz and Underwood as responses. Four sets of eight syllables were selected from the 0, 33, 67, and 100% Glaze lists, with the additional criterion that the values approximate Archer's (1960) revaluation. Within each set similarity was minimized and four different combinations of S-R pairings were used to control for associative effects. Both T_L and T_R were either 2 sec. or 4 sec. In other respects the experiment replicated that of Jantz and Underwood (1958).

Results.—As found in the earlier study, total syllables correct during the recall trial displayed significant heterogeneity of variance between groups. The total letters correct, a criterion used by Feldman and Underwood (1957), indicated homogeneity of variance. Consequently, the latter measure served in the analysis of variance of treatment effects reported here. However, comparable results were obtained throughout for the heterogeneous total syllable scores.

The overall means and *SD*s for AV were 10.30 and 4.77, 14.70 and 5.29, 16.60 and 4.16, and 19.40 and 3.82 for 0, 33, 67, and 100%, respectively. A trend analysis for these means revealed a highly significant linear trend ($F = 22.34$, $P < .001$). The overall means and *SD*s for T_L were 12.52 and 5.17 for 2 sec. and 18.20 and 3.92 for 4 sec. The corresponding T_R statistics were 13.90 and 4.54 for 2 sec. and 16.83 and 4.55 for 4 sec. The main effects for AV, T_L, and T_R were all significant ($F = 10.85$, $df = 3/64$, $P < .001$; $F = 26.98$, $df = 1/64$, $P < .001$; $F = 7.14$, $df = 1/64$, $P < .01$). Of the interactions, only the $T_L \times T_R$ approached significance ($F = 2.85$, $df = 1/64$, $P < .10$). This trend suggested that T_R had little effect on the 2-sec. T_L groups but did on the 4-sec. T_L groups. That is, the 4-sec. recall condition was most effective in combination with the 4-sec. S-R learning condition.

The present results appear to verify the assumption made by Jantz and Underwood (1958) that their reversal for 0% and 33% represented a sampling artifact. The results also indicate that R-S learning is sensitive to exposure duration during both the learning and recall periods.

REFERENCES

ARCHER, E. J. A re-evaluation of the meaningfulness of all possible CVC trigrams. *Psychol. Monogr.*, 1960, **74**(10, Whole No. 497).

FELDMAN, S. M., & UNDERWOOD, B. J. Stimulus recall following paired-associate learning. *J. exp. Psychol.*, 1957, **53**, 11–15.

JANTZ, E. M., & UNDERWOOD, B. J. R-S learning as a function of meaningfulness and degree of S-R learning. *J. exp. Psychol.*, 1958, **56**, 174–179.

KAUSLER, D. H., & TRAPP, E. P. The effects of position and exposure duration on irrelevant cue learning. Paper read at Southwestern Psychological Association, Little Rock, Arkansas, April 1961.

POSTMAN, L., ADAMS, P. A., & PHILLIPS, L. W. Studies in incidental learning: II. The effects of association value and of the method of testing. *J. exp. Psychol.*, 1955, **49**, 1–10.

ROSENBERG, S. Exposure interval in incidental learning. *Psychol. Rep.*, 1959, **5**, 675.

(Received May 20, 1961)

[1] This study is based on a thesis submitted by the first author to the Graduate School, St. Louis University in partial fulfillment for the Master of Arts degree. Portions of this paper were presented at the 1961 meeting of the Midwestern Psychological Association.

Journal of Experimental Psychology
1962, Vol. 64, No. 1, 95–96

SUPPLEMENTARY REPORT: EFFECTS OF INSTRUCTIONS ON EXTINCTION AND RECOVERY OF A CONDITIONED AVOIDANCE RESPONSE

K. E. MOYER AND RICHARD H. LINDLEY[1]

Carnegie Institute of Technology *Trinity University, Texas*

Lindley and Moyer (1961) found that, in the case of a classically conditioned finger withdrawal response, informing *S*s just prior to extinction training that the UCS would no longer be delivered produced rapid extinction. The present study replicated that study, but with a conditioned avoidance procedure, and in addition tested an implication of the drive level (*D*) interpretation of their result; viz., that the appropriate instructions could raise *D* level and lead to an increase in conditioned responding.

Method.—The same experimental room, apparatus, intertrial stimulus intervals, method of adjusting shock level, and pretests with the tone alone were used (Lindley & Moyer, 1961). The CS was a .5-sec. tone which preceded a 1.5-sec. electric shock without overlap. The finger could only move up when the CS and UCS were presented. No shock was delivered if *S* responded to the CS and kept his finger raised ¼ in. or more during the UCS interval. If *S* did not avoid the shock, moving the finger up ¼ in. escaped the shock.

The booth in which *S* sat was dark. The initial instructions were similar to the previous instructions except that the avoidance rather than the classical CR procedure was explained. The *S* was told that we wanted to condition automatic movements of the finger but that he should not try to prevent his finger from moving.

A factorial design involving number of acquisition conditioning trials (to four CRs in 5 consecutive trials, or that plus 20 trials), instructions prior to extinction (Neutral or Inhibitory), and instructions after 20 extinction trials (No Instructions or Resume UCS) was used. The Neutral instructions prior to extinction were: "Be sure to let your finger jump up to the tone when it feels like it"; the Inhibitory instructions were: "There will be no more shock presented from now on. I want you to try to prevent your finger from moving when the tone is presented." After 20 extinction trials half the *S*s were given an additional 5 extinction trials without any instructions (No Instructions); the remaining *S*s were told: "From now on the tone will be followed by shock on most of the trials. Remember to let your finger jump up to the tone if it feels like it" (Rseume UCS). The extinction procedure was continued for 5 trials (i.e., no shock was delivered). The normal intertrial interval was used between Trials 20 and 21 for the No Instructions groups; this interval was longer for the Resume UCS groups owing to the reading of the instructions.

The 56 *S*s (both men and women) from Carnegie Institute of Technology who reached criterion within 35 trials were assigned to the eight groups (N per group = 7) by a randomized block procedure. Twenty-seven *S*s were rejected owing to failure to reach criterion within 35 trials, apparatus failure, or unwillingness to be shocked.

Results.—The mean number of trials to reach the criterion of four CRs in five consecutive trials was 10.98, $SD = 7.16$. There were no significant differences among the groups in reaching this criterion ($F = 0.30$).

In extinction, a CR was defined as any response that occurred during the .5-sec.

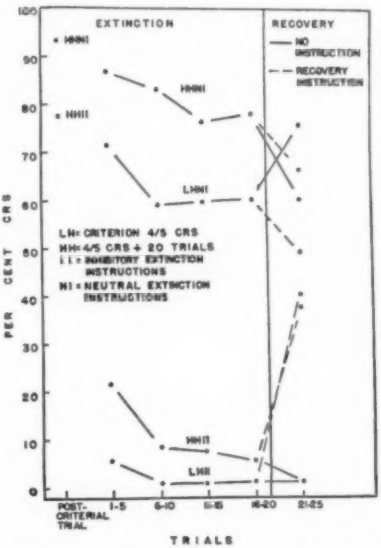

FIG. 1. Percentage of CRs for blocks of five trials during extinction and recovery training and the percentage of CRs on the first postcriterial trial for the 4/5 + 20 trial groups.

[1] We would like to thank James H. Korn for collecting the data for this study.

duration of the CS. The variances of the number of CRs in extinction were highly heterogeneous ($\chi^2=62.53$, $df=3$, $P<.001$). Owing to the heterogeneity of variance, the data of the groups that received neutral instructions and of the groups that received inhibitory instructions were analyzed separately and the χ^2 median test was used at times instead of the more usual t or F tests.

A χ^2 median test showed that the inhibitory instructions significantly reduced the number of CRs elicited in the 20 extinction trials as compared to the neutral instructions ($\chi^2 = 23.17$, $df = 1$, $P < .001$). Figure 1 shows the extinction and recovery data for all groups and also the percentage of CRs on the first trial after *S*s reached the $\frac{8}{8}$ criterion in the two groups with the $\frac{8}{8}$ criterion plus 20 trials. The groups with the additional 20 conditioning trials gave more responses in extinction than the groups that merely met the $\frac{8}{8}$ criterion (for the two groups with Neutral instructions, $t = 1.27$, $df = 26$, $P > .05$; for the groups with Inhibitory instructions, $t = 2.18$, $df = 26$, $P < .05$).

The instructions designed to raise *D* level after 20 extinction trials had a significant effect on the groups that received Inhibitory instructions in extinction (median test $\chi^2 = 8.58$, $df = 1$, $P < .01$). The same instructions had no demonstrable effect on the groups that received Neutral instructions in extinction ($\chi^2 = 1.29$, $df = 1$, $P > .05$).

As in classical conditioning, inhibitory instructions dramatically reduced the number of CRs during extinction after avoidance conditioning, especially after the larger number of conditioning trials. This fits in with the notion that a habit factor is built up through reinforcement and is independent of a performance or drive factor which can be manipulated by instructions. The results also indicate that, for *S*s who have been given inhibitory instructions, the appropriate instructions can increase the number of CRs after extinction, presumably by raising *D* level. The latter result was not found for *S*s who have been given neutral instructions in extinction, perhaps due to the already high level of responding of these groups after the 20 extinction trials.

Reference

Lindley, R. H., & Moyer, K. E. Effects of instructions on the extinction of a conditioned finger-withdrawal response. *J. exp. Psychol.*, 1961, **61**, 82–88.

(Received May 18, 1961)